Maastricht University

The effects of spatial resolution and physiological contrast on fMRI patterns

Citation for published version (APA):

Gardumi, A. (2017). *The effects of spatial resolution and physiological contrast on fMRI patterns*. [Doctoral Thesis, Maastricht University]. Gildeprint Drukkerijen. https://doi.org/10.26481/dis.20170324ag

Document status and date:
Published: 01/01/2017

DOI:
10.26481/dis.20170324ag

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 19 Apr. 2024

The effects of spatial resolution and physiological contrast on fMRI patterns

Anna Gardumi

© Anna Gardumi, Maastricht 2017

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the publisher.

The work in this thesis was supported by the Marie Curie Initial Training Network (MC-ITN) “Methods in Neuroimaging” (NeuroPhysics).

Cover	Anna Gardumi
Production	Gildeprint
ISBN	978-94-6233-590-5

The effects of spatial resolution and physiological contrast on fMRI patterns

DISSERTATION

to obtain the degree of Doctor at Maastricht University,
on the authority of Rector Magnificus, Prof. Dr. R.M. Letschert,
in accordance with the decision of the Board of Deans,
to be defended in public on Friday 24 March 2017 at 14:00 hours

by

Anna Gardumi

Supervisor
Prof. Dr. Elia Formisano

Co-supervisor
Dr. Kâmil Uludağ

Assessment Committee
Prof. Dr. Rainer Goebel (Chair)
Prof. Dr. Klaus Scheffler (Max Planck Institute for Biological Cybernetics, Tübingen)
Dr. Marcus Herdener (University of Zurich, Zürich)
Dr. Federico De Martino

Contents

1 General introduction **9**

1.1 FMRI: state-of-the-art . . . 12

1.1.1 Limitations of "standard" fMRI techniques . . . 13

1.1.2 Arterial Spin Labeling . . . 15

The ASL principle . . . 15

(Dis)advantages of ASL versus BOLD fMRI . . . 16

Implementation techniques . . . 17

1.1.3 Ultra-high field fMRI . . . 19

The special case of ASL at ultra-high field . . . 20

1.1.4 Multivariate pattern analysis . . . 21

Preprocessing of functional data for MVPA . . . 22

Feature selection . . . 23

Feature extraction . . . 23

Training . . . 24

Testing and score . . . 24

1.2 Auditory cortex: state-of-the-art . . . 25

1.2.1 Tonotopy in auditory cortex . . . 27

1.2.2 Speech and voice processing . . . 28

1.3 Outline of the thesis . . . 30

2 The effect of spatial resolution on decoding accuracy in fMRI MVPA **37**

2.1 Abstract . . . 38

2.2 Introduction . . . 39

2.3 Material and methods . . . 41

2.3.1 Subjects . . . 41

2.3.2 Stimuli and task . . . 41

2.3.3 Data acquisition . . . 42

2.3.4 Data analysis: preprocessing and univariate analysis . . . 43

2.3.5 Data analysis: multivariate analysis . . . 43
Feature selection . . . 44
Feature extraction . . . 45
Classification . . . 45
Statistical testing . . . 46
2.3.6 Effect of spatial smoothing . . . 46
2.3.7 Reconstruction of data at lower effective spatial resolution . 46
2.3.8 Influence of tSNR, CNR and head motion on MVPA . . . 49
2.3.9 Voxel-wise tissue type ranking . . . 49
2.4 Results . . . 50
2.4.1 Univariate analysis . . . 50
2.4.2 Vowel and speaker decoding from the original data . . . 51
2.4.3 Effects of spatial resolution and smoothing on decoding . . 53
2.4.4 Effect of spatial resolution on decoding . . . 54
2.4.5 Influence of tSNR, CNR, and head motion on MVPA . . . 56
2.4.6 Influence of tissue type on decoding . . . 56
2.5 Discussion . . . 57
2.6 Supplementary material 1 . . . 66
2.6.1 Changing resolution by cropping the k-space and its influence on decoding . . . 66
2.6.2 Control analyses for resolution and smoothing effects . . . 66
2.6.3 Validation of the BV likelihood method . . . 69
2.7 Supplementary material 2 . . . 71
2.7.1 Statistical testing by permutation at group level . . . 71

3 Tonotopic maps in human auditory cortex using ASL 75
3.1 Abstract . . . 76
3.2 Introduction . . . 77
3.3 Material and methods . . . 79
3.3.1 Subjects . . . 79
3.3.2 Stimulus design and presentation . . . 79
3.3.3 MRI acquisition . . . 80
3.3.4 Preprocessing . . . 81
3.3.5 BOLD and CBF time courses . . . 83
3.3.6 Computation of the BOLD- and CBF-based tonotopic maps 84
3.3.7 Comparison of BOLD- vs CBF-based tonotopic maps . . . 85

3.3.8 Baseline CBF quantification . . . 86
3.3.9 Voice selective regions . . . 87
3.3.10 Vein masks from susceptibility weighted images . . . 87
3.4 Results . . . 88
3.4.1 Activation in auditory cortex . . . 88
3.4.2 CBF and BOLD tonotopies . . . 90
3.4.3 Perfusion map and parcellation of the primary auditory cortex . . . 91
3.4.4 Voice selective regions . . . 93
3.4.5 Vein masks from susceptibility weighted images . . . 95
3.5 Discussion . . . 95
3.5.1 Limitations and benefits of tonotopy using ASL . . . 99
3.6 Conclusions . . . 100
3.7 Supporting Information . . . 107
3.7.1 Vein masks . . . 107
3.7.2 Power spectrum of 3T pCASL with 2D single-shot readout . 107
3.7.3 Activation in auditory cortex . . . 107
3.7.4 CBF quantification . . . 107
3.7.5 Anatomically defined and baseline CBF based PAC . . . 108
3.7.6 Voice selective regions . . . 108
3.7.7 Voice selective regions for a large range of initial vertex-level threshold . . . 112

4 Comparison of 3T and 7T ASL for concurrent CBF and BOLD imaging 117
4.1 Abstract . . . 118
4.2 Introduction . . . 119
4.3 Materials and Methods . . . 122
4.3.1 Subjects . . . 122
4.3.2 Data Acquisition . . . 122
4.3.3 ASL sequences . . . 122
4.3.4 Preprocessing . . . 126
4.3.5 Perfusion calculation . . . 127
4.3.6 SNR measures . . . 128
4.3.7 CBF and BOLD activation . . . 129
4.3.8 Perfusion territories . . . 130
4.4 Results . . . 131

4.4.1 Perfusion maps . . . 131
4.4.2 SNR measures . . . 133
Temporal SNR . . . 133
Perfusion SNR . . . 135
4.4.3 CBF and BOLD activation . . . 135
4.4.4 Perfusion territory CBF . . . 138
4.5 Discussion . . . 140
4.5.1 The influence of magnetic field strength . . . 140
Differences in absolute CBF values between 3T and 7T . . . 141
4.5.2 The effect of spatial resolution . . . 142
4.5.3 Dual-echo vs single-echo ASL . . . 143
4.5.4 The influence of labelling scheme . . . 144
4.5.5 The effect of background suppression and readout scheme . 145
4.6 Conclusions . . . 147
4.7 Supplementary Material . . . 156
4.7.1 Single-subject and group values of the mean GM CBF . . . 156
4.7.2 Mean GM perfusion in the anatomically defined GM mask . 158
4.7.3 Creation of perfusion territory masks . . . 160
4.7.4 SNR measures for 2^{nd} echo data . . . 160
4.7.5 Activation and baseline CBF for 2^{nd} echo data . . . 162

5 Summary & General discussion **165**
5.1 Summary . . . 167
5.2 Conclusions and outlook . . . 170

Knowledge valorization **177**

Acknowledgments **183**

Curriculum Vitae **191**

Publications **195**

Chapter 1

General introduction

How much do sounds affect our daily life? Weekdays start with the repetitive tone of the alarm going off, weekend days with a pleasant tweeting of the birds outside your window. The rough gurgle coming from the kitchen is telling you that coffee is brewed and the bright ring of the toaster is the final invitation for breakfast. While you are biking to work, the rattling sound of the chain reminds you that it's time for some maintenance and a car horn can warn you from an unseen danger. After a long day, you are finally back home listening to some relaxing music when the doorbell suddenly rings. Answering to the intercom, a voice says "It's me!" and you immediately recognize your best friend voice: "Hi Joy, come on in!"

This brief collection of daily moments illustrates how much impact sounds have on our lives. It also powerfully shows how sounds are promptly and automatically analyzed by our brain. Somehow, we are able to appreciate not only differences in loudness and pitch, but also to infer more complex information such as, for instance, the identity of the person who is talking to us from the other side of the intercom. *Hearing* (or *auditory perception* or *audition*) is traditionally defined as one of the five senses and, witnessing its ecological relevance, it has been object of study for centuries in the medical context of hearing loss cases. Recently, non-invasive neuroimaging techniques have allowed studying the neural correlates of auditory mechanisms in both the healthy and diseased brain. However, still today, little is known about the complex processing that the auditory system performs in order to extract relevant acoustic features and form meaningful sound representations.

Similarly, many other cognitive neuroscience questions have been addressed using neuroimaging and (functional) magnetic resonance imaging (fMRI) has become one of the most important methods to study the living brain. Besides its non-invasiveness, (f)MRI is attractive for both clinical and research purposes due to the wide range of structural to functional information that it can provide. (F)MRI data are gathered with a continuously increasing spatial and temporal resolution, sensitivity, specificity and robustness, combined with other experimental methods, and analyzed with a variety of processing methods in the attempt to grasp the complexity of the brain and its cognitive processes. Despite the outstanding progress done in the past years, fundamental questions still remain concerning the biophysical mechanisms contributing to the (f)MRI signal itself and their implications for the analysis and interpretation of the results.

The present thesis employed and combined advanced fMRI methods, namely

arterial spin labeling, ultra-high field fMRI and multivariate pattern analysis, to shed further light on the physiological underpinnings of the methods themselves and their application in the context of auditory research. This introductory chapter will present the state-of-the-art of fMRI methods and auditory research focusing on the aspects relevant for the projects included in this thesis.

1.1 FMRI: state-of-the-art

FMRI is a non-invasive neuroimaging technique and represents one of the core methods used in cognitive neuroscience research. FMRI based on the blood oxygenation level-dependent (BOLD) contrast is currently the predominant method used for activation studies (Logothetis, 2008; Bandettini, 2009) as it provides (indirect) measure of neuronal activity with (relatively) high spatial and temporal resolution (Logothetis et al., 2001). An increase in neuronal activity causes a localized change in cerebral metabolic rate of oxygen ($CMRO_2$), cerebral blood flow (CBF) and cerebral blood volume (CBV). Consequently, the concentration ratio of oxygenated and deoxygenated hemoglobin changes and, due to their different magnetic properties (dia- and paramagnetic, respectively), gives rise to the measurable BOLD effect (Ogawa et al., 1992). BOLD signal reflects, therefore, the combined effect of several physiological changes (namely, $CMRO_2$, CBF and CBV), which are linked to neuronal activity through neurovascular coupling (see Figure 1.1).

The typical hemodynamic response function (HRF) is characterized by an elusive initial small dip, which might or might not be present, a rapid increase of the signal which peaks around 4-8 s followed by a slower decrease and, in most cases, an undershoot before returning to baseline. Although there is a wide variability in HRF shape (latency, amplitude, width, initial dip, undershoot, etc.) between different brain areas, subjects, subject populations and experimental tasks, the overall behavior is highly consistent (Handwerker et al., 2012) (see Figure 1.2). Therefore, in most standard activation fMRI studies, the HRF shape is assumed fixed and is modeled with the so called *canonical HRF* resulting by the convolution of a box-car function (representing the stimulus) with a double gamma function (Friston et al., 1998). This model is used within a General Linear Model (GLM) analysis to statistically infer activation voxel-by-voxel. Finally, functional contrast maps (t-statistics) are then used to evaluate activation differences across

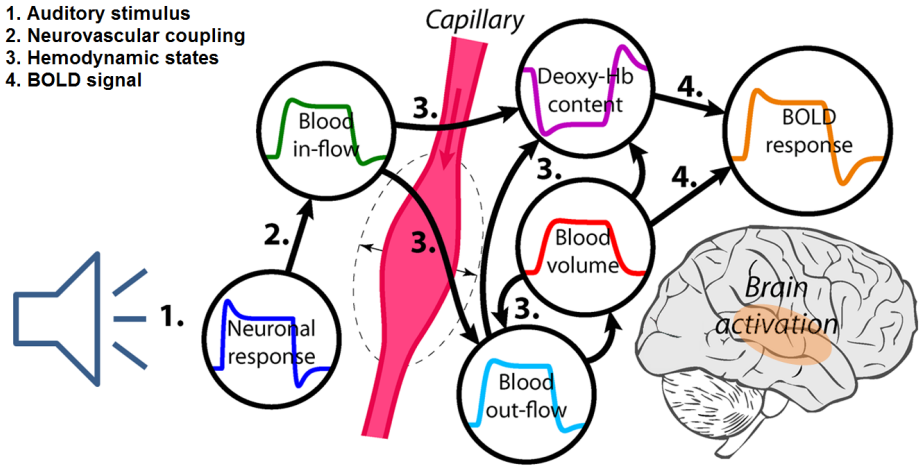

FIGURE 1.1: Schematic illustration of the many physiological changes that follow an external (auditory) stimulus and lead to the BOLD response usually measured in fMRI. Adapted from Havlicek et al. (2015).

different experimental conditions. Such voxel-by-voxel analysis is the most used analysis technique in task-based fMRI studies and is also known as *univariate* or *mass-univariate* analysis. Because of the large number of voxels in a brain (typically on the order of hundreds of thousands of voxels in typical 3T fMRI datasets), multiple comparison corrections (e.g., Bonferroni, false discovery rate, cluster thresholding, etc.) are necessary in order to account for false positives at the global level.

1.1.1 Limitations of "standard" fMRI techniques

Due to the nature of the fMRI signal, the ultimate limit of both temporal and spatial resolution is imposed by the temporal and spatial extent of the hemodynamic response. However, spatial and temporal resolutions are further limited by the hardware (scanners and coils) and software implementations (pulse sequences, reconstruction algorithms) used. Currently, standard fMRI studies performed at 3T MR scanners using gradient-echo echoplanar imaging (GE-EPI), are characterized by a temporal resolution of 2-3 s and a voxel size around 3x3x3 mm^3. Such a relatively large voxel size is necessary to guarantee enough signal-to-noise ratio (SNR) while maintaining a reasonable temporal resolution and brain coverage. Nevertheless, considering that the cerebral cortex has a thickness of 1.0-4.5 mm

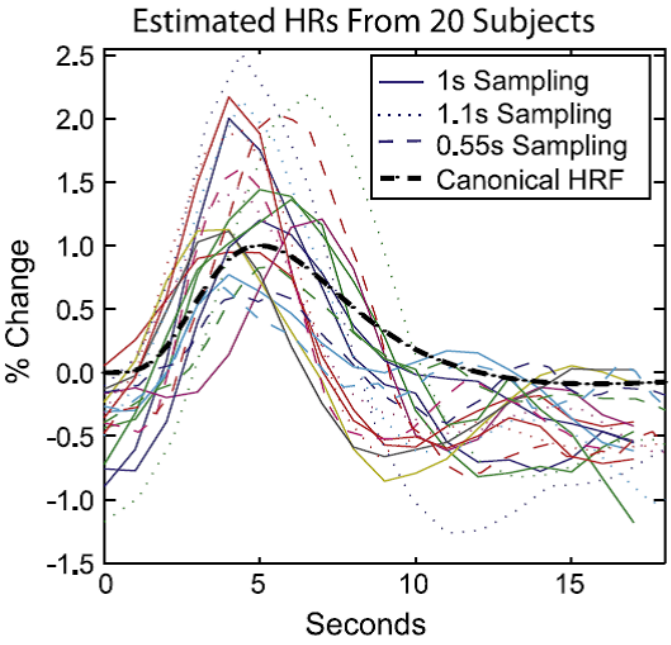

FIGURE 1.2: Hemodynamic responses from 20 subjects averaged across a ROI in primary sensorimotor cortex in response to a single button press. From Handwerker et al. (2012).

(Fischl and Dale, 2000), the measured signal is affected by partial volume effects between gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF).

Additionally, although changes in the BOLD signal should ideally be localized at the site of neuronal activity, changes in the oxy-deoxyhemoglobin ratio and blood volume propagates through the vasculature giving rise to poorly spatially specific signal at the level of draining veins. In the case of GE-EPI, the BOLD signal is dominated by draining veins contribution, thus its spatial specificity is degraded (Olman et al., 2007; Ugurbil et al., 2003; Uludag et al., 2009).

Finally, the complex nature of the BOLD signal, which results by the combined effect of $CMRO_2$, CBF and CBV, complicates the interpretation of its changes. BOLD signal is measured in arbitrary units and its absolute values are influenced by many different factors, dependent both on physiological and acquisition parameters. Thus, BOLD baseline shifts are usually ignored and only percent signal changes are considered.

The effort of overcoming these limitations and opening the possibilities for new research questions, has, therefore, driven the fMRI community to the development of new technologies and methods. Among these, a selection composed by non-BOLD techniques (arterial spin labeling), ultra-high field fMRI, and advanced statistical analyses (multivariate pattern analysis) will be described in the next paragraphs.

1.1.2 Arterial Spin Labeling

Arterial Spin Labeling (ASL) is a non-invasive fMRI technique that measures cerebral blood flow (CBF) using arterial water as an endogenous tracer (Borogovac and Asllani, 2012).

The ASL principle

The proton spins of the arterial water are inverted (or saturated) by a radio-frequency (RF) pulse in a region proximal to the region we want to image (i.e., the entire brain or a part of it; see Figure 1.3). A time delay (typically between 700 ms and 2000 ms) is introduced in the sequence scheme to allow such inverted spins to flow through the arterial tree till the imaging slab. After this time interval, a so-called *label* or *tag image* is acquired. Moreover, for each label image a *control image* is acquired, for which no inversion (or saturation) RF pulse is applied to the

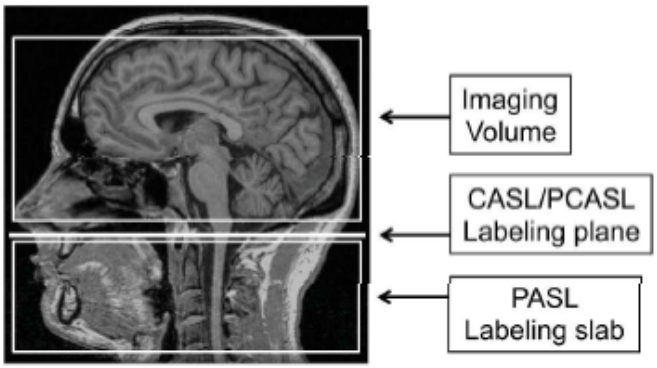

FIGURE 1.3: Scheme illustrating the ASL acquisition geometry. The labeling plane or slab represents the location where the magnetization of the arterial water protons is inverted (termed *magnetically labeled*). The imaging volume might be the entire brain (as depicted here) or a part of it. From Alsop et al. (2015).

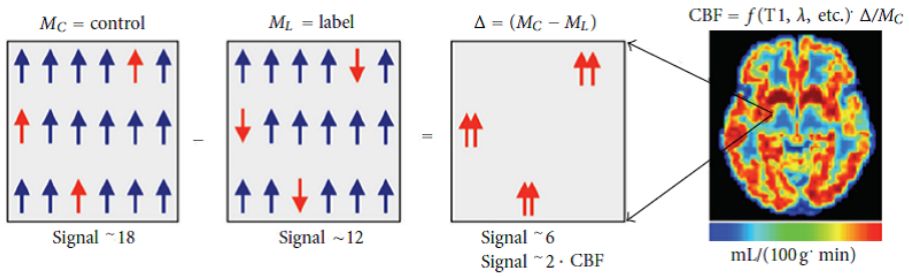

FIGURE 1.4: Schematic representation of the principle of ASL signal. The subtraction of the label from the control image yields the CBF of the imaging volume. CBF values are expressed in ml/100g/min and are higher in gray matter compared to white matter. From Borogovac and Asllani (2012).

spins. The difference between control and label image is therefore proportional to the blood delivered into each voxel (see Figure 1.4). A calibration image together with a model of physiological and MR parameters allow calculating at voxel level the quantitative value of CBF in ml/100g/min (typical physiological units).

(Dis)advantages of ASL versus BOLD fMRI

Measuring CBF offers some important advantages with respect to the BOLD contrast. First, CBF is quantitative and physiologically meaningful, while BOLD is measured in arbitrary units and is the complex result of interplaying variations in CBF, CBV and $CMRO_2$. Thus, CBF is more suitable for longitudinal studies and its absolute values can be directly used for inter-session and inter-centers comparisons. The interpretation of CBF changes are straightforward, whilst more complex for BOLD (i.e., percent signal) changes. Second, CBF changes are more closely related to neuronal activity and therefore better spatially localized and more spatially specific than BOLD changes whose spatial specificity is degraded by draining vein bias (Aguirre et al., 2005). Third, CBF power spectrum is essentially frequency independent because perfusion is calculated as a pairwise comparison of consecutive temporal points. CBF is therefore suitable for detecting slow changes (e.g., emotional changes, drug therapies, ...), while BOLD contrast suffers from temporal autocorrelation and is not suitable for detecting changes that are more than 90s apart (Aguirre et al., 2002). Finally, typical ASL techniques allow the simultaneous measure of both CBF and BOLD contrast. The latter is indeed obtained from the same ASL fMRI time series by averaging (instead of

subtracting) label-control pairs. Besides the evident advantage of acquiring simultaneous and already co-registered CBF and BOLD time series for comparison purposes, a recent study has tried to exploit this characteristic to improve the estimation of the hemodynamic response to stimuli which cannot be easily repeated or averaged (Simon et al., 2013).

Some challenging limitations of current ASL techniques are low SNR compared to BOLD techniques, relatively slow temporal resolution (due to both the time delay necessary for the spins to reach the imaging slab and the need of acquiring label-control pairs), transit time confounds (i.e., differences in blood transit time between brain areas (MacIntosh et al., 2010; Mildner et al., 2014)), and partial volume effects (relatively large voxels are needed to compensate for the low SNR). In the following section, describing the different ASL implementations, some other and sequence specific drawbacks will be pointed out.

Implementation techniques

Based on the same principle, several ASL implementations were developed differing in how spin labeling is achieved and how artifacts, such as magnetization transfer (MT) effects, are controlled. Historically, the two major categories of arterial spin labeling are: 1) *Continuous ASL* (CASL) in which a long (a few seconds) labeling pulse is applied on a thin labeling plane; 2) *Pulsed ASL* (PASL) in which a short (a few ms) labeling pulse is applied on an extended (several cm) labeling slab (see Figure 1.3). Moreover, within each category, several variants are possible and, more recently, a third category has been developed: *pseudo-continuous ASL* (pCASL), which tries to profit from the benefits of both continuous and pulsed ASL.

Continuous ASL

In CASL techniques, spin labeling is achieved by adiabatic inversion, which requires a long labeling pulse. This requirement has two major disadvantages. Firstly, a long off-resonance RF pulse causes signal loss in label images due to MT effect. In contrast, control images do not require a labeling pulse and therefore are not affected by MT. Thus, label and control images differ not only because of the inflowing of magnetically inverted blood but also due to the MT effect. The estimation of CBF is therefore confounded. This problem was reduced by the introduction of an amplitude-modulated RF pulse for the control images inducing a double inversion (i.e., no net magnetization) with a MT similar to the labeling one

(Alsop and Detre, 1998). Secondly, long pulses are very demanding on the hardware and most often not supported from commercially available MR scanners. In these cases, the use of a second coil applying the inversion pulse becomes necessary (Zhang et al., 1995; Zaharchuk et al., 1999). Such dual-coil techniques allow solving both major disadvantages of CASL, however, it is difficult to implement and requires extra-hardware.

Pulsed ASL

PASL techniques use short RF pulses (usually 10-15 ms), thus they suffer less from the MT effect, however, at the cost of lower SNR (typically 30-50% lower SNR than CASL). One of the most used PASL scheme is the flow-sensitive alternating inversion recovery (FAIR) sequence by (Kim, 1995; Schwarzbauer et al., 1996). Both label and control images are achieved by inversion recovery scan: the label image with a slice-selective inversion and the control image with a non-slice-selective inversion. The difference between the two scans is proportional to the amount of inflowing blood because only in the case of slice-selective inversion "new" uninverted blood can flow into the imaging slab. One of the greatest advantages of FAIR is, therefore, its easy implementation, although its extension to multi-slice capability was not trivial due to imperfections in the slice profile (Kwong et al., 1995; Golay et al., 2004).

PASL schemes are therefore advantageous because of the absence of MT effects, no need of extra-hardware, high labeling efficiency and their (relatively) easy implementation. However, they offer lower SNR (with respect to CASL schemes) and they are sensitive to transit time delays and slice profile artifacts.

Pseudo-continuous ASL

PCASL is a technique (Silva and Kim, 1999; Dai et al., 2008) which was developed to combine the advantages of both continuous and pulsed approaches, namely the high SNR from CASL and the high labeling efficiency without requirement of extra hardware from PASL (Wu et al., 2007). Such performances are achieved by labeling the blood in the feeding arteries via flow-driven adiabatic inversion: equally spaced selective Hanning shape RF pulses and rapid repeated gradients result in a continuous labeling with high efficiency. Albeit the labeling duration is quite long (typically $\sim$1.8 s), the absence of MT effects is guaranteed because the same RF pulses scheme is used for both labeling and control images. The latter, however, is characterized by an alternating sign of the RF pulses and a zero mean average gradient between each pair of pulses, which cause no net magnetization

whilst equivalent RF power, spatially equal to the labeling condition. Moreover, pCASL was demonstrated to outperform CASL and PASL in term of reliability in repeated measures (Xu et al., 2010; Gevers et al., 2011). Because of all these advantages, pCASL was chosen by the ASL community as the recommended ASL acquisition scheme for clinical applications (Alsop et al., 2015).

Nevertheless, it is worth noting that in the context of functional studies it is still not clear which is the preferable ASL technique (especially between pCASL and PASL schemes) because of the critical importance of short TR acquisition. While pCASL sampling rate is limited by the necessity of (relatively) long labeling duration (1.8 s labeling duration is recommended, resulting in a usual TR of $\sim$4-5 s (Alsop et al., 2015), PASL schemes can offer shorter TRs (up to $\sim$2 s) and therefore higher sampling rate. The method of choice, therefore, is determined by a trade-off between sampling rate and SNR, and its optimization depends on the specific study. In the fourth chapter of this thesis, we will address this issue comparing different ASL schemes and their performances in the context of functional studies.

1.1.3 Ultra-high field fMRI

The term ultra-high field MRI commonly refers to field strengths larger than 4T and, nowadays, in most cases of human MRI studies, it refers to 7T MRI.

Ultra-high field strengths provide the opportunity for higher spatial resolution and specificity of the BOLD signal. The higher SNR, due to the stronger static magnetic field and thus to the increased susceptibility difference between oxy- and deoxy-hemoglobin, allows higher spatial resolution ($\sim$0.7-1.0 mm isotropic) and, therefore, lower partial volume effects. Thanks to voxels sizes below 1 mm isotropic, it is now possible to investigate the organization of the cortex in layers and columns (Yacoub et al., 2008; Polimeni et al., 2010). Moreover, the SNR benefit allows to employ low SNR techniques such as spin echo (SE) sequences, which were demonstrated having higher spatial specificity to neuronal activity due to dominant contribution of the signal originating from the micro-vasculature (Yacoub et al., 2003; Uludag et al., 2009).

While ultra-high field fMRI offers the benefits of increased spatial SNR and BOLD contrast, smaller voxel sizes, reduced partial-volume effects, and increased sensitivity to microvasculature, it also presents several challenges. With the increase of the magnetic field strength the field inhomogeneities (both B_0 and B_1)

become more severe and their correction more critical. Macroscopic B_0 inhomogeneities (i.e., inhomogeneity of the static field) are mainly caused by tissue/air interfaces (Truong et al., 2006) and result in signal loss and geometrical distortions. Moreover, B_1 inhomogeneities result in variable flip angles within the imaging slab and, therefore, in locally variable SNR and signal contrast. Because of the increased susceptibility effects, the sensitivity to head motion is also increased. Higher magnetic field strength also means need of higher frequency for RF pulses and, thus, higher energy deposition in the tissue. The specific absorption rate (SAR) is, therefore, a more stringent constraint in ultra-high field than in lower field MRI. Finally, the imaged volume being equal, higher spatial resolution translates into larger number of voxels and, therefore, slower performance in data (pre)processing, more severe multiple comparison problem in univariate analyses and higher dimensionality in multivariate analyses.

Many efforts have been made in the last years to improve field homogeneity, correct image artifacts and optimize (pre)processing algorithms and, today, it is commonly accepted that 7T BOLD fMRI outperforms 3T BOLD fMRI. This is indeed witnessed by the rapid increase of the number of 7T fMRI studies as well as by the growing number of imaging center acquiring 7T MR scanners.

The special case of ASL at ultra-high field

The benefit of measuring ASL at higher field strength is two-fold: with the increase of the static magnetic field strength 1) the overall SNR increases (SNR $\sim$ B_0, as true for BOLD techniques as well) and 2) the longitudinal relaxation time (T_1) increases ($T_1 \sim B_0^{1/3}$) leading to a slower decay of the labeling. One would, thus, expect that 7T is the field of choice for ASL research. However, there is not yet sufficient evidence of 7T ASL outperforming 3T ASL and part of the ASL community is skeptical about ultra-high field ASL (e.g., Teeuwisse et al., 2010). The reason of such skepticism is that the above-mentioned challenges of ultra-high field MRI are particularly concerning in the case of ASL techniques. B_0 inhomogeneities due to susceptibility effects are present at the level of both the imaging volume and the labeling plane/slab (i.e., the proximal part of the brain or neck where the labeling pulse is applied). Moreover, RF pulse delivery efficiency rapidly drops off toward the lower part of the head coil, i.e., again in a critical region for ASL labeling. For these reasons, the labeling efficiency is importantly reduced and actual flip angles are reduced by at least 50% with respect to the

nominal flip angles (Teeuwisse et al., 2010). One possibility to compensate for low B_1 efficiency would be using higher RF pulse amplitude, but this is practically not feasible because some ASL schemes are already close to the upper limit of SAR. For these reasons, the FAIR scheme seems to be the most promising ASL sequence at 7T as it is a pulsed scheme with relatively low SAR demand and the labeling slab is centered on the imaging slab, i.e. at level of the brain (and not of the neck), where the RF coil has high B_1^+ efficiency (Gardener et al., 2009).

In the fourth chapter of this thesis, we will also address this open question comparing the performances of the FAIR scheme implemented at 7T and at 3T. Moreover, the comparison will be extended to other PASL and pCASL implementations at 3T.

1.1.4 Multivariate pattern analysis

In the recent years, machine learning and pattern recognition algorithms have been increasingly employed in neuroimaging studies under the name of *multivariate pattern analysis* (MVPA), *multi-voxel pattern analysis*, *brain decoding* or *brain reading*. In contrast to univariate analysis (where the fMRI signal of each voxel is analyzed independently), MVPA is based on the simultaneous analysis of the fMRI signal measured from all voxels in the brain (or in pre-selected ROI(s)). Building such a multi-dimensional space, the algorithm is able to *learn* a functional relationship between activation patterns and cognitive states or perceptual representations. The use of MVPA instead of univariate analysis represents, therefore, a switch from activation-based to information-based analysis (Kriegeskorte and Bandettini, 2007). Furthermore, MVPA allows overcoming the idea of a necessarily spatially localized activation by allowing for detection of spread (non-local) activation patterns. Finally, by considering all voxels simultaneously in a multi-dimensional space, it implicitly solves the multiple comparison problem typical for standard univariate analysis. Importantly, MVPA was demonstrated to be able to detect subtle effects even in absence of detectable univariate effects leading, hence, to the recent great interest of the neuroimaging community in this technique. However, little is known about the physical and physiological underpinnings of MVPA in fMRI. In the second chapter of this thesis, we tackled this issue by combining MVPA and high-resolution fMRI at ultra-high field.

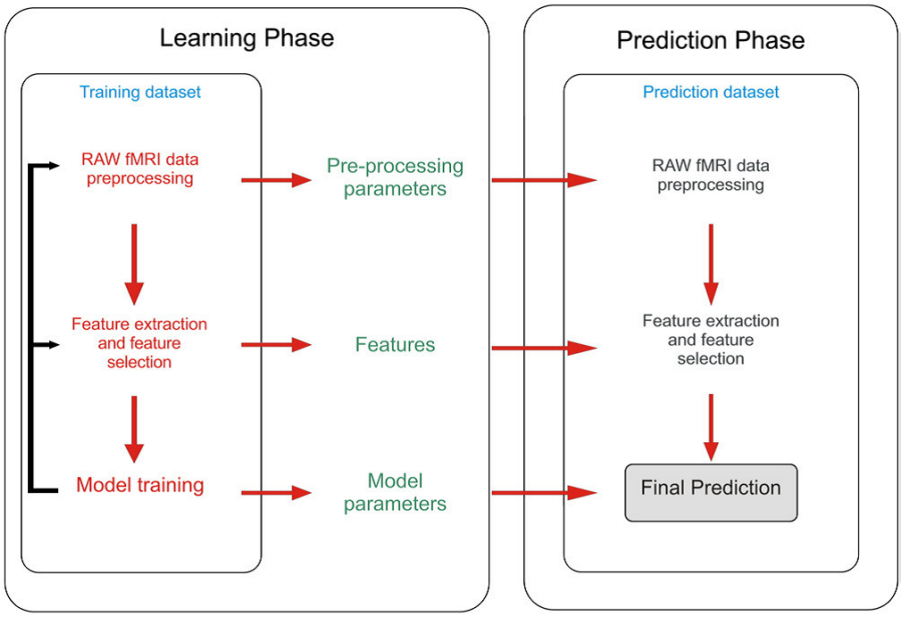

FIGURE 1.5: Scheme of the main steps in a generic pattern recognition algorithm as used in fMRI analysis. From Formisano et al. (2008b).

Figure 1.5 schematically illustrates the steps generally performed in machine learning algorithms applied to fMRI data. In the following paragraphs, each of these steps will be briefly described.

Preprocessing of functional data for MVPA

FMRI experiments usually consist of several functional runs acquired while participants are performing a certain task. Within the same experiment session one high-resolution anatomical image is also acquired. Standard preprocessing includes slice time correction, motion correction of the functional runs and coregistration between functional and anatomical images. Transformation to a standardized space (e.g., MNI space) is usually avoided unless decoding is performed across subjects, but this latter case is not considered here. High-pass temporal filtering is applied in order to increase SNR and remove slow "scanner drifts" from the fMRI signal. Spatial smoothing might or might not be applied (in Chapter 2 of this thesis we will investigate this choice).

Feature selection

In machine learning, *feature selection* means to select those features in the data that are most relevant for the predictive model one wishes the algorithm to learn. In the particular case of fMRI MVPA, the features to consider are the voxels. One straightforward approach would be considering all the brain voxels. However, this choice is usually suboptimal because it results in an extremely high dimensional space (the number of voxels in a brain is on the order of hundred thousand voxels or more), and including voxels which are not informative for the predictive model under question is known to degrade MVPA performance. Moreover, restricting the large number of features is particularly important in fMRI MVPA because the number of observations (examples) is usually very limited (in the order of tens of observations), a problem called *curse of dimensionality* (i.e., number of observations $\ll$ number of features). If a priori knowledge is available, it is possible to select features on the basis of ROIs. Otherwise, both univariate and multivariate selection approaches have been proposed. Univariate approaches consist in selecting voxels that show stimulus related BOLD responses or univariate contrast effects (e.g., as assessed by t-statistics). These selection methods offer the advantage of simplifying the interpretation of the MVPA results. However, they do not consider the multivariate nature of the model that the algorithm is trying to learn and may, thus, result in suboptimal feature selection. Multivariate approaches were, therefore, introduced to take into account the multivariate nature of the problem already during the voxel selection step. One potential drawback of such approaches is that they are highly time consuming and suffer from high computational complexity (for more details about multivariate approaches, see Guyon and Elisseeff, 2003; Formisano et al., 2008b, and references therein).

Feature extraction

In machine learning, the *feature extraction* step consists in transforming the input data (which could be also text or images) of each feature in a numerical value, which can be handled by the algorithm. In fMRI MVPA, the feature extraction step consists in summarizing each trial of each voxel with a single numerical value. A common procedure is to perform a GLM with a predictor for each class of stimuli and for each trial and extract the resulting β- or t-values as features for the classifier. Alternatively, the mean or the maximum amplitude of the HRF response of each trial have been used as features. The choice of the summarizing

variable depends strongly on the design of the experiment (e.g., mean amplitude of the HRF response is more suitable for block designs than for event related designs), but it remains arbitrary and there is not yet definitive consensus on the optimal choice.

Training

Once the feature selection and extraction are performed, the resulting feature set is used as input data for *training* the machine learning algorithm. In case of supervised learning, each training example is completed by a label defining its class (cognitive state). The algorithm can therefore learn the relationship between feature patterns and cognitive states of interest. The output of this learning process is a weight value for each feature such that the weight of highly informative voxels is maximized and that of irrelevant ones is minimized. The weight vector obtained allows to divide the multidimensional space by means of hyperplanes, which best differentiate the classes of interest.

Testing and score

In order to evaluate whether the machine learning algorithm has correctly learned to classify the different classes (cognitive states), it is necessary to *test* the predictive model with a new (unseen) dataset of examples. The test dataset undergoes the same preprocessing, feature selection and extraction performed on the training dataset. Finally, the learned model is applied to the new examples and the predicted labels are compared with the true (unseen by the algorithm) labels. The accuracy of these predictions (i.e., the percentage of correct label prediction) is generally used as the performance measure of the machine learning algorithm. Note that, in order to be a valid test, the dataset used for testing must be completely independent from the training dataset. In most cases, the test dataset is represented by a functional run acquired during the same session and not used in the training of the algorithm, or by an ensemble of single trials left out from the runs used for the training. While in the first case there might only be a suspicion of optimistic bias, in the latter case the assumption of independence does not hold. Therefore, it is important to test the significance of the results through an unbiased statistical test: the *permutation test*. The permutation test is a non-parametric statistical test, which provides an unbiased estimate of the chance level without requiring independence of the training and testing dataset. Under

the null hypothesis, there is no class information in the data. Thus, the labels can be randomly permuted without altering the expected accuracy. The test consists in randomly permuting the labels of the training and/or testing dataset many times and using such permuted labels to train and test the classifier. Repeating this operation many times yields an unbiased estimate of the null distribution. The p-value of the "true" decoding corresponds to the fraction of occurrences for which the accuracy with the permuted labels is higher than the original one. The permutation test offers, therefore, the important advantage of providing an unbiased estimate of the chance level, but at the expense of long computation times.

In this section *FMRI: state-of-the-art*, we have illustrated the methodological framework of the studies included in the present thesis. Neuroimaging and analysis methods here described were applied to investigate the human auditory cortex. The following section will briefly introduce the current neuroscientific knowledge on sound processing and auditory cortex.

1.2 Auditory cortex: state-of-the-art

In everyday life, we are able to discriminate different sounds by effortlessly processing their acoustic properties. One main property of sounds is frequency and previous studies have shown that the processing of frequencies happens all along the auditory system from the cochlea till the auditory cortex (AC; see Figure 1.6). The first processing step starts already in the ears where sounds, i.e. a pressure wave travelling through the air, are transformed into vibrations and, finally, into electrical impulses by the cochlea. This spiral-shaped organ is composed by several components. Among these, the basilar membrane (BM) has an important role in sound processing. BM progressively varies in thickness and stiffness along the cochlea and, therefore, resonates at different location depending on the frequency of the incoming sound. In this way, the cochlea spatially segregates the frequency components of incoming sounds: high frequencies are represented near its base and low frequencies near its apex. This spatially ordered representation of frequencies is called *tonotopy* (from Greek *tono* = frequency and *topos* = place) and is a functional organization which is preserved at later stages of the auditory pathways till the auditory cortex.

In addition to such frequency analysis, more complex analyses of the sounds

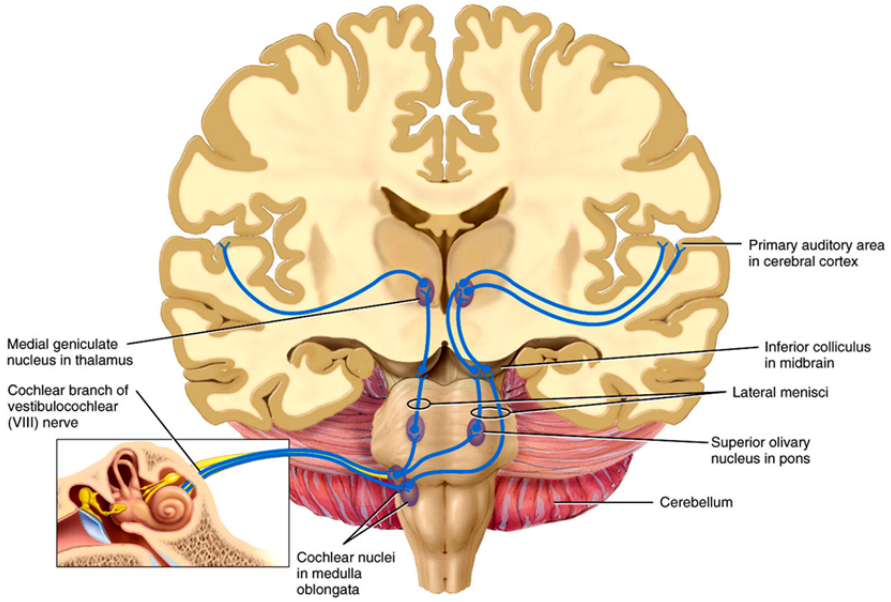

FIGURE 1.6: Cochlea and auditory pathway. Sounds reach the inner ear as air pressure waves and are transformed into mechanical vibration and finally in neural impulses by the cochlea. The neural signal travels through the vestibulocochlear nerve to the cochlear nuclei, the superior olivary complex, the inferior colliculus in the midbrain, the medial geniculate nucleus, and finally the auditory cortex. The resulting sound information is therefore the result of the processing at each stage of the auditory pathway. From Tortora and Derrickson (2008).

are performed at the cortex level. It is indeed commonly accepted that complex sounds are processed in a hierarchical way from low- to high-level features, but there is no consensus yet on the specific acoustic features used and the location of each processing stage. In analogy with the visual system, an auditory "what/where model" has been proposed, in which the "what" corresponds to processes in a temporal pathway (antero-ventral stream) and the "where" to those in a parietal pathway (postero-dorsal stream) (Rauschecker, 1998; Romanski et al., 1999; Rauschecker and Tian, 2000). Other studies suggested that the "where" pathway may be responsible for spectral motion (Belin and Zatorre, 2000) and action related (Warren et al., 2005) processing, therefore the terms "how" and "do" stream were proposed instead.

In this thesis, we will focus on low-level sound processing, namely tonotopy

(in Chapter 3), and high-level sound processing, namely speech and voice (in Chapter 2), both at the level of the auditory cortex. Thus, the following two paragraphs will briefly describe these two topics.

1.2.1 Tonotopy in auditory cortex

Imaging studies have investigated the representation of frequencies in AC and have consistently found a map of preferred frequencies both in animals (Merzenich and Brugge, 1973; Merzenich et al., 1975; Reale and Imig, 1980) and in humans (Formisano et al., 2003; Humphries et al., 2010; Da Costa et al., 2011; Moerel et al., 2012). The fundamental characteristic of the tonotopic map in human auditory cortex is generally described as two main gradients going from high-to-low and from low-to-high preferred frequency. This reversal gradient is located on Heschl's gyrus (HG) and its surrounding areas in planum temporale (PT) and planum polare (PP). However, there is not yet consensus on the extension and the orientation of the gradients with respect to their correspondence with the core and belt areas of AC (see Figure 1.7, Panel B and C). One interpretation, called *classical*, locates the auditory core along HG in agreement with cytoarchitectonic evidences (1.5T: Seifritz et al. (2006); 3T: Schonwiesner et al. (2002); 7T: Formisano et al. (2003)). Another interpretation, called *orthogonal*, locates the core across HG including also the neighboring areas in PP and PT and bases its foundations in across-species consistency as the long axis of monkey auditory core runs parallel to the superior temporal gyrus (STG) (1.5 T: Talavage et al. (2004); Woods et al. (2010); 3 T: Humphries et al. (2010); Striem-Amit et al. (2011); Langers and van Dijk (2012); 7 T: Da Costa et al. (2011)).

Exploring auditory cortex outside its primary core, several other tonotopic gradients are generally observed (see Figure 1.7, Panel D), although they are rarely interpreted because highly variable across subjects. Nevertheless, they might provide useful information for functional parcellation of the auditory cortex and additional studies are needed to better understand their topological and functional properties (Moerel et al., 2014).

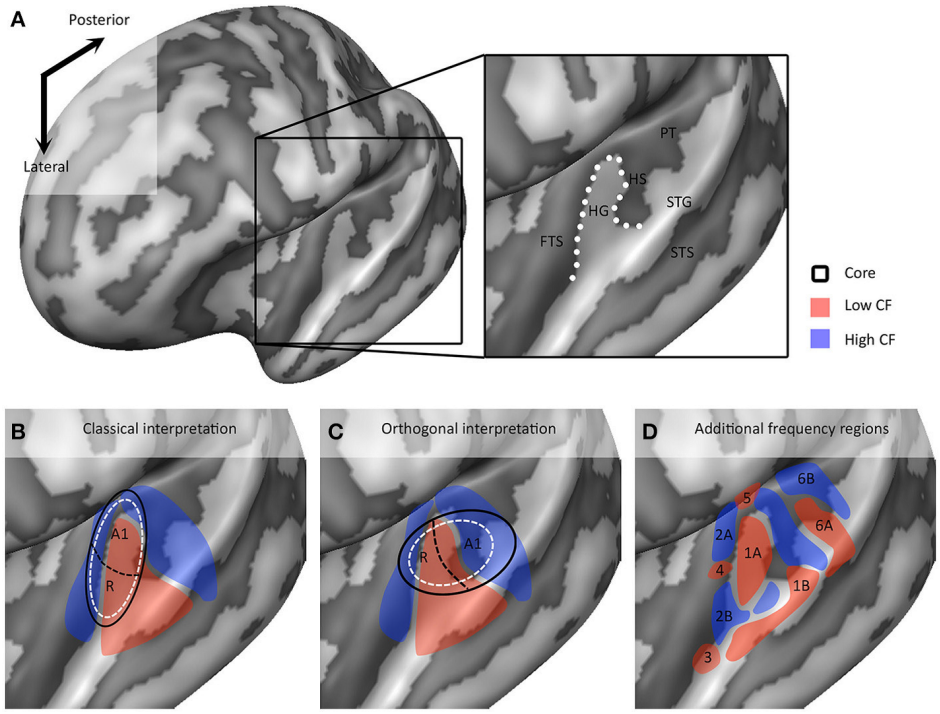

FIGURE 1.7: Tonotopy in human auditory cortex. Panel A shows the anatomical location and components of the auditory cortex on an inflated representation of the left hemisphere: first transverse sulcus (FTS), Heschl's gyrus and sulcus (HG, HS), planum temporale (PT), planum polare (PP), superior temporal gyrus and sulcus (STG, STS). Panel B and C illustrate the classical and orthogonal interpretation of tonotopic maps with primary auditory cortex running along or orthogonal HG, respectively. Panel D depicts additional tonotopic gradients that may be found in AC. From Moerel et al. (2014).

1.2.2 Speech and voice processing

Proceeding along the temporal pathway, studies have shown an increasing complexity of preferred stimuli and, therefore, a hierarchical processing from low-level characteristics (e.g., loudness, pitch, ...) to more abstract categories (e.g., animal cries, tool noises, ...). Speech is a particular case of the abstract categories and its investigation is ecologically of particular relevance. FMRI studies have reliably found that, in certain temporal areas, the BOLD response to vocal human sounds is larger than for other sound categories. These areas consist in three clusters located in STG and STS posterior to HS, lateral portion of HG/HS, and (mainly in the right hemisphere) anteriorly along STG/STS. There are different

interpretations of this finding. One possible interpretation is that these regions are highly functionally specialized and selective for human voices. Hence, they have been termed *voice areas* (Belin et al., 2000). An alternative interpretation is that human voices are particular rich in specific low-level acoustic features that happen to be processed in that specific area of the cortex. Supporting this interpretation is the finding that auditory cortex was comparably activated (i.e., no significant univariate contrast was found) in a study where human and non-human sounds (female voice, cat and guitar) were finely matched in terms of acoustical and physical properties (duration, root mean square power, temporal profile, ...) (Staeren et al., 2009). Interestingly, information about the sound categories, as assessed through multivariate analysis (please see the section 1.1.4 *Multivariate pattern analysis* for details about this type of statistical analysis), was present and spread over the entire auditory cortex, including early areas commonly attributed to low-level processing. These findings strengthened those by (Formisano et al., 2008a) who showed the involvement of early auditory regions, in addition to commonly called high-level regions, in the abstract representation of speech content (vowels) and speaker identity. Importantly, these studies demonstrated the presence of relevant decoding information despite the absence of an univariate effect.

1.3 Outline of the thesis

In Chapter 2, multivariate pattern analysis was combined with ultra-high field fMRI and reconstruction of the data at different spatial resolutions in order to investigate the effect of spatial resolution and smoothing on decoding of speech content and speaker identity from the auditory cortex. It was shown that the high spatial frequencies are informative for vowel decoding, while the informative pattern for speaker identity resides at lower spatial frequencies. These findings suggest that the optimal spatial resolution of acquisition and preprocessing scheme for MVPA depend on the decoding task and brain area investigated.

In Chapter 3, the tonotopic organization of the human auditory cortex was investigated using arterial spin labeling fMRI. The feasibility of CBF-based tonotopy was demonstrated for the first time and its good agreement with BOLD-based tonotopy represents a reciprocal validation. In addition, primary auditory cortex was shown to be characterized by higher baseline perfusion with respect to the rest of the cortex. We, therefore, proposed baseline perfusion as an additional biomarker for primary auditory cortex and its parcellation.

In Chapter 4, different ASL MRI approaches implemented at different spatial resolutions and different field strengths (3T and 7T) were compared for baseline perfusion mapping and functional purposes. In the case of low resolution and no use of parallel imaging, 3T ASL implementations yielded the highest SNR and are therefore recommended over 7T PASL. In contrast, in the case of high resolution and/or use of parallel imaging, 7T PASL is preferable over 3T ASL.

In Chapter 5, the findings of the empirical chapters of this thesis were summarized and discussed. On the basis of the outcomes and limitations of our studies, we speculate about future directions of investigation and possible methods allowing such explorations.

Bibliography

G. K. Aguirre, J. A. Detre, E. Zarahn, and D. C. Alsop. Experimental design and the relative sensitivity of BOLD and perfusion fMRI. *Neuroimage*, 15(3):488–500, 2002.

G. K. Aguirre, J. A. Detre, and J. Wang. Perfusion fMRI for functional neuroimaging. *Int Rev Neurobiol*, 66:213–36, 2005.

D. C. Alsop and J. A. Detre. Multisection cerebral blood flow MR imaging with continuous arterial spin labeling. *Radiology*, 208(2):410–6, 1998.

D. C. Alsop, J. A. Detre, X. Golay, M. Gunther, J. Hendrikse, L. Hernandez-Garcia, H. Lu, B. J. MacIntosh, L. M. Parkes, M. Smits, M. J. van Osch, D. J. Wang, E. C. Wong, and G. Zaharchuk. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. *Magn Reson Med*, 73(1):spcone, 2015.

P. A. Bandettini. What's new in neuroimaging methods? *Ann N Y Acad Sci*, 1156: 260–93, 2009.

P. Belin and R. J. Zatorre. 'What', 'where' and 'how' in auditory cortex. *Nat Neurosci*, 3(10):965–6, 2000.

P. Belin, R. J. Zatorre, P. Lafaille, P. Ahad, and B. Pike. Voice-selective areas in human auditory cortex. *Nature*, 403(6767):309–12, 2000.

A. Borogovac and I. Asllani. Arterial Spin Labeling (ASL) fMRI: advantages, theoretical constrains, and experimental challenges in neurosciences. *Int J Biomed Imaging*, 2012:818456, 2012.

S. Da Costa, W. van der Zwaag, J. P. Marques, R. S. Frackowiak, S. Clarke, and M. Saenz. Human primary auditory cortex follows the shape of Heschl's gyrus. *J Neurosci*, 31(40):14067–75, 2011.

W. Dai, D. Garcia, C. de Bazelaire, and D. C. Alsop. Continuous flow-driven inversion for arterial spin labeling using pulsed radio frequency and gradient fields. *Magn Reson Med*, 60(6):1488–97, 2008.

B. Fischl and A. M. Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci U S A*, 97(20):11050–5, 2000.

E. Formisano, D. S. Kim, F. Di Salle, P. F. van de Moortele, K. Ugurbil, and R. Goebel. Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, 40(4):859–69, 2003.

E. Formisano, F. De Martino, M. Bonte, and R. Goebel. "Who" is saying "what"? brain-based decoding of human voice and speech. *Science*, 322(5903):970–3, 2008a.

E. Formisano, F. De Martino, and G. Valente. Multivariate analysis of fMRI time series: classification and regression of brain responses using machine learning. *Magn Reson Imaging*, 26(7):921–34, 2008b.

K. J. Friston, P. Fletcher, O. Josephs, A. Holmes, M. D. Rugg, and R. Turner. Event-related fMRI: characterizing differential responses. *Neuroimage*, 7(1):30–40, 1998.

A. G. Gardener, P. A. Gowland, and S. T. Francis. Implementation of quantitative perfusion imaging using pulsed arterial spin labeling at ultra-high field. *Magn Reson Med*, 61(4):874–82, 2009.

S. Gevers, M. J. van Osch, R. P. Bokkers, D. A. Kies, W. M. Teeuwisse, C. B. Majoie, J. Hendrikse, and A. J. Nederveen. Intra- and multicenter reproducibility of pulsed, continuous and pseudo-continuous arterial spin labeling methods for measuring cerebral perfusion. *J Cereb Blood Flow Metab*, 31(8):1706–15, 2011.

X. Golay, J. Hendrikse, and T. C. Lim. Perfusion imaging using arterial spin labeling. *Top Magn Reson Imaging*, 15(1):10–27, 2004.

I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3:1157–1182, 2003.

D. A. Handwerker, J. Gonzalez-Castillo, M. D'Esposito, and P. A. Bandettini. The continuing challenge of understanding and modeling hemodynamic variation in fMRI. *Neuroimage*, 62(2):1017–23, 2012.

M. Havlicek, A. Roebroeck, K. Friston, A. Gardumi, D. Ivanov, and K. Uludag. Physiologically informed dynamic causal modeling of fMRI data. *NeuroImage*, 122:355 – 372, 2015.

C. Humphries, E. Liebenthal, and J. R. Binder. Tonotopic organization of human auditory cortex. *Neuroimage*, 50(3):1202–11, 2010.

S. G. Kim. Quantification of relative cerebral blood flow change by flow-sensitive alternating inversion recovery (FAIR) technique: application to functional mapping. *Magn Reson Med*, 34(3):293–301, 1995.

N. Kriegeskorte and P. Bandettini. Analyzing for information, not activation, to exploit high-resolution fMRI. *Neuroimage*, 38(4):649–62, 2007.

K. K. Kwong, D. A. Chesler, R. M. Weisskoff, K. M. Donahue, T. L. Davis, L. Ostergaard, T. A. Campbell, and B. R. Rosen. MR perfusion studies with T1-weighted echo planar imaging. *Magn Reson Med*, 34(6):878–87, 1995.

D. R. Langers and P. van Dijk. Mapping the tonotopic organization in human auditory cortex with minimally salient acoustic stimulation. *Cereb Cortex*, 22 (9):2024–38, 2012.

N. K. Logothetis. What we can do and what we cannot do with fMRI. *Nature*, 453 (7197):869–78, 2008.

N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843):150–7, 2001.

B. J. MacIntosh, N. Filippini, M. A. Chappell, M. W. Woolrich, C. E. Mackay, and P. Jezzard. Assessment of arterial arrival times derived from multiple inversion time pulsed arterial spin labeling MRI. *Magn Reson Med*, 63(3):641–7, 2010.

M. M. Merzenich and J. F. Brugge. Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Res*, 50(2):275–96, 1973.

M. M. Merzenich, P. L. Knight, and G. L. Roth. Representation of cochlea within primary auditory cortex in the cat. *J Neurophysiol*, 38(2):231–49, 1975.

T. Mildner, K. Muller, S. Hetzer, R. Trampel, W. Driesel, and H. E. Moller. Mapping of arterial transit time by intravascular signal selection. *NMR Biomed*, 27 (5):594–609, 2014.

M. Moerel, F. De Martino, and E. Formisano. Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. *J Neurosci*, 32(41):14205–16, 2012.

M. Moerel, F. De Martino, and E. Formisano. An anatomical and functional topography of human auditory cortical areas. *Front Neurosci*, 8:225, 2014.

S. Ogawa, D. W. Tank, R. Menon, J. M. Ellermann, S. G. Kim, H. Merkle, and K. Ugurbil. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci U S A*, 89(13):5951–5, 1992.

C. A. Olman, S. Inati, and D. J. Heeger. The effect of large veins on spatial localization with GE BOLD at 3 T: Displacement, not blurring. *Neuroimage*, 34(3): 1126–35, 2007.

J. R. Polimeni, B. Fischl, D. N. Greve, and L. L. Wald. Laminar analysis of 7T BOLD using an imposed spatial activation pattern in human V1. *Neuroimage*, 52(4):1334–46, 2010.

J. P. Rauschecker. Cortical processing of complex sounds. *Curr Opin Neurobiol*, 8 (4):516–21, 1998.

J. P. Rauschecker and B. Tian. Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc Natl Acad Sci U S A*, 97(22):11800–6, 2000.

R. A. Reale and T. J. Imig. Tonotopic organization in auditory cortex of the cat. *J Comp Neurol*, 192(2):265–91, 1980.

L. M. Romanski, B. Tian, J. Fritz, M. Mishkin, P. S. Goldman-Rakic, and J. P. Rauschecker. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci*, 2(12):1131–6, 1999.

M. Schonwiesner, D. Y. von Cramon, and R. Rubsamen. Is it tonotopy after all? *Neuroimage*, 17(3):1144–61, 2002.

C. Schwarzbauer, S. P. Morrissey, and A. Haase. Quantitative magnetic resonance imaging of perfusion using magnetic labeling of water proton spins within the detection slice. *Magn Reson Med*, 35(4):540–6, 1996.

E. Seifritz, F. Di Salle, F. Esposito, M. Herdener, J. G. Neuhoff, and K. Scheffler. Enhancing BOLD response in the auditory system by neurophysiologically tuned fMRI sequence. *Neuroimage*, 29(3):1013–22, 2006.

A. C. Silva and S. G. Kim. Pseudo-continuous arterial spin labeling technique for measuring CBF dynamics with high temporal resolution. *Magn Reson Med*, 42 (3):425–9, 1999.

A. B. Simon, V. E. M. Griffeth, E. C. Wong, and R. B. Buxton. A novel method of combining blood oxygenation and blood flow sensitive magnetic resonance imaging techniques to measure the cerebral blood flow and oxygen metabolism responses to an unknown neural stimulus. *PloS one*, 8(1):e54816, 2013.

N. Staeren, H. Renvall, F. De Martino, R. Goebel, and E. Formisano. Sound categories are represented as distributed patterns in the human auditory cortex. *Curr Biol*, 19(6):498–502, 2009.

E. Striem-Amit, U. Hertz, and A. Amedi. Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding fMRI. *PLoS One*, 6(3):e17832, 2011.

T. M. Talavage, M. I. Sereno, J. R. Melcher, P. J. Ledden, B. R. Rosen, and A. M. Dale. Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *J Neurophysiol*, 91(3):1282–96, 2004.

W. M. Teeuwisse, A. G. Webb, and M. J. P. van Osch. Arterial spin labeling at ultra-high field: All that glitters is not gold. *International Journal of Imaging Systems and Technology*, 20(1):62–70, 2010.

G. J. Tortora and B. H. Derrickson. *Principles of anatomy and physiology*. Wiley, 12th edition edition, 2008.

T. K. Truong, D. W. Chakeres, D. Q. Beversdorf, D. W. Scharre, and P. Schmalbrock. Effects of static and radiofrequency magnetic field inhomogeneity in ultra-high field magnetic resonance imaging. *Magn Reson Imaging*, 24(2):103–12, 2006.

K. Ugurbil, L. Toth, and D. S. Kim. How accurate is magnetic resonance imaging of brain function? *Trends Neurosci*, 26(2):108–14, 2003.

K. Uludag, B. Muller-Bierl, and K. Ugurbil. An integrative model for neuronal activity-induced signal changes for gradient and spin echo functional imaging. *Neuroimage*, 48(1):150–65, 2009.

J. E. Warren, R. J. Wise, and J. D. Warren. Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends Neurosci*, 28(12):636–43, 2005.

D. L. Woods, T. J. Herron, A. D. Cate, E. W. Yund, G. C. Stecker, T. Rinne, and X. Kang. Functional properties of human auditory cortical fields. *Front Syst Neurosci*, 4:155, 2010.

W. C. Wu, M. Fernandez-Seara, J. A. Detre, F. W. Wehrli, and J. Wang. A theoretical and experimental investigation of the tagging efficiency of pseudocontinuous arterial spin labeling. *Magn Reson Med*, 58(5):1020–7, 2007.

G. Xu, H. A. Rowley, G. Wu, D. C. Alsop, A. Shankaranarayanan, M. Dowling, B. T. Christian, T. R. Oakes, and S. C. Johnson. Reliability and precision of pseudo-continuous arterial spin labeling perfusion MRI on 3.0 T and comparison with 15o-water PET in elderly subjects at risk for Alzheimer's disease. *NMR Biomed*, 23(3):286–93, 2010.

E. Yacoub, T. Q. Duong, P. F. Van De Moortele, M. Lindquist, G. Adriany, S. G. Kim, K. Ugurbil, and X. Hu. Spin-echo fMRI in humans using high spatial resolutions and high magnetic fields. *Magn Reson Med*, 49(4):655–64, 2003.

E. Yacoub, N. Harel, and K. Ugurbil. High-field fMRI unveils orientation columns in humans. *Proc Natl Acad Sci U S A*, 105(30):10607–12, 2008.

G. Zaharchuk, P. J. Ledden, K. K. Kwong, T. G. Reese, B. R. Rosen, and L. L. Wald. Multislice perfusion and perfusion territory imaging in humans with separate label and image coils. *Magn Reson Med*, 41(6):1093–8, 1999.

W. Zhang, A. C. Silva, D. S. Williams, and A. P. Koretsky. NMR measurement of perfusion using arterial spin labeling without saturation of macromolecular spins. *Magn Reson Med*, 33(3):370–6, 1995.

Chapter 2

The effect of spatial resolution on decoding accuracy in fMRI MVPA

Based on:

Gardumi, A., Ivanov, D., Hausfeld, L., Valente, G., Formisano, E., and Uludağ, K. (2016). *The effect of spatial resolution on decoding accuracy in fMRI multivariate pattern analysis*. NeuroImage, 132, 32-42

2.1 Abstract

Multivariate pattern analysis (MVPA) in fMRI has been used to extract information from distributed cortical activation patterns, which may go undetected in conventional univariate analysis. However, little is known about the physical and physiological underpinnings of MVPA in fMRI as well as about the effect of spatial smoothing on its performance. Several studies have addressed these issues, but their investigation was limited to the visual cortex at 3T with conflicting results. Here, we used ultra-high field (7T) fMRI to investigate the effect of spatial resolution and smoothing on decoding of speech content (vowels) and speaker identity from auditory cortical responses. To that end, we acquired high-resolution (1.1 mm isotropic) fMRI data and additionally reconstructed them at 2.2 and 3.3 mm in-plane spatial resolutions from the original k-space data. Furthermore, the data at each resolution were spatially smoothed with different 3D Gaussian kernel sizes (i.e., no smoothing or 1.1, 2.2, 3.3, 4.4, or 8.8 mm kernels). For all spatial resolutions and smoothing kernels, we demonstrate the feasibility of decoding speech content (vowel) and speaker identity at 7T using Support Vector Machine (SVM) MVPA. In addition, we found that high spatial frequencies are informative for vowel decoding and that the relative contribution of high and low spatial frequencies is different across the two decoding tasks. Moderate smoothing (up to 2.2 mm) improved the accuracies for both decoding of vowels and speakers, possibly due to reduction of noise (e.g., residual motion artifacts or instrument noise) whilst still preserving information at high spatial frequency. In summary, our results show that - even with the same stimuli and within the same brain areas - the optimal spatial resolution for MVPA in fMRI depends on the specific decoding task of interest.

2.2 Introduction

Functional magnetic resonance imaging (fMRI) is currently the most popular non-invasive method to investigate human brain structure and function. It indirectly measures neural activity primarily via the blood oxygenation level-dependent (BOLD) effect. Standard univariate statistical analysis (i.e., general linear model (GLM) analysis) of the task-based fMRI data has been utilized to detect voxel-wise differences of BOLD activation levels and, thus, to infer which brain areas are involved in a certain task. In recent years, multivariate pattern analysis (MVPA) has been used in fMRI to extract information from spatially distributed activation patterns, which may go undetected in conventional univariate analysis. Reliable decoding of information from fMRI data acquired at 3T has been demonstrated from activation patterns in different brain areas (Haxby et al., 2001; Cox and Savoy, 2003; Haynes and Rees, 2005; Kamitani and Tong, 2005; Kriegeskorte and Bandettini, 2007; Formisano et al., 2008). Different biophysical hypotheses have been proposed to explain the ability of MVPA on fMRI data to detect information inaccessible with GLM. It has been suggested that MVPA is sensitive to information encoded at the sub-millimeter scale of neuronal functional columns. Such information, even if sampled at the lower resolution of standard fMRI voxel sizes (e.g., 3x3x3 mm^3), may be accessible by MVPA due to local variations and irregularities in the columnar organization, resulting in weak but consistent biases in fMRI responses of the different voxels (Boynton, 2005; Kamitani and Tong, 2005; Haynes and Rees, 2006; Kamitani and Tong, 2006; Kriegeskorte and Bandettini, 2007); this mechanism is, therefore, named *hyperacuity* or *voxel biased sampling*. Alternatively, the transposition from high spatial frequency components of columns preferences to lower spatial frequency of the fMRI signal may be attributed to the cortical vasculature. This hypothesis is based on the fact that, using the standard gradient echo (GE) MRI sequences, the fMRI signal stems mostly from veins draining blood from a given tissue volume (see Uludag et al., 2009). Thus, a specific vein could be more sensitive to one neuronal population than another introducing a spatial bias. Hence, this hypothesis is known as *biased draining regions* (Kamitani and Tong, 2005; Gardner et al., 2006; Kamitani and Tong, 2006; Kriegeskorte and Bandettini, 2007; Gardner, 2010; Kamitani and Sawahata, 2010; Kriegeskorte et al., 2010; Shmuel et al., 2010).

According to another hypothesis, MVPA may rely on large spatial scale non-columnar organization (Op de Beeck, 2010), such as radial preference maps (Freeman et al., 2011). Since MVPA represents a computational scheme to non-locally average the fMRI signal, in this framework, MVPA would be able to detect low spatial frequency information too weak to be detected with univariate analysis.

Note that these hypotheses are not mutually exclusive (see Shmuel et al., 2010; Swisher et al., 2010). Nevertheless, they do predict testable effects of spatial smoothing on decoding performance. Op de Beeck has shown that spatial smoothing does not deteriorate decoding performance of objects and orientations from activation patterns in lateral occipital cortex and V1, respectively (Op de Beeck, 2010). He interpreted these results as an argument against hyperacuity and in favor of large-scale organization. Further support for this hypothesis comes from the finding that it is possible to decode across experimental sessions performed in different days (Freeman et al., 2011). In contrast, several studies (Swisher et al., 2010; Alink et al., 2013; Misaki et al., 2013) demonstrated that spatial smoothing decreases decoding accuracies for orientation and ocular dominance from V1 data, suggesting relevant information content at the individual voxel level. The few investigations so far on the underlying mechanisms of MVPA on fMRI data and the effect of spatial smoothing have been limited to the early visual cortex. In addition, they have been restricted to a small set of stimuli and decoding tasks (e.g., decoding of orientation, ocular dominance and direction of motion) and have yielded conflicting evidence.

The main goal of the current study is to investigate how information at different spatial resolutions contributes to MVPA decoding. We employed ultra-high field (7T) fMRI to acquire high-resolution data (1.1 mm isotropic), which were then reconstructed at different effective spatial resolutions from original k-space data to evaluate the effects of spatial resolution on MVPA decoding performance. Based on an experimental paradigm and on stimuli that were used in a previous fMRI study at 3 T (Formisano et al., 2008), we presented speech stimuli (vowels) from different speakers and considered the single-trial decoding of vowels and speakers from auditory cortical response patterns. Compared to conventional 3T fMRI, 7T fMRI presents several advantages, such as higher signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR), and therefore the possibility of higher spatial resolution with lower partial volume effects and greater spatial specificity (Yacoub et al., 2005; Uludag et al., 2009; Polimeni et al., 2010). On the other hand, it presents challenges such as larger distortions, sensitivity to motion,

and larger number of voxels to be handled by the decoding algorithm (Formisano and Kriegeskorte, 2012). Therefore, we also investigated the effects of temporal SNR, CNR, head motion, and of typical noise-reduction steps (spatial smoothing) on MVPA performances.

2.3 Material and methods

2.3.1 Subjects

Ten healthy volunteers (seven females, age range 25–32) with normal hearing took part in this experiment. Informed consent was obtained from all participants according to the approval by the Ethical Committee of the Faculty of Psychology and Neuroscience, University of Maastricht.

2.3.2 Stimuli and task

We used the same auditory stimuli as in the study by Formisano et al. (2008) consisting of three vowels (/a/, /i/, /u/) spoken by three different speakers (sp1: female, sp2: male, sp3: male). For each of these 9 conditions, three different tokens were included in order to introduce acoustic variability. All stimuli were equated in length to 230 ms and in sound intensity by matching their root mean square amplitude. For more details about the stimulus properties, please see (Formisano et al., 2008). Prior to the functional experiment, participants were familiarized with the stimuli and were able to recognize the corresponding vowels and speakers. During the fMRI experiment, subjects were instructed to attentively listen to the stimuli while fixating a white cross in the center of the screen. The stimuli were presented in the silent gap between two subsequent image acquisitions (see below).

In order to ensure the engagement of the participants in both listening and fixating tasks, the participants performed a one back-task on the speaker dimension irrespectively of the spoken vowel: 10% of the total number of trials were catch-trials (signaled to the participants by the fixation cross turning red for 100 ms), in which the subjects were asked to report whether the speaker of the last heard sound was the same as the previous one. Subjects performed the task by pressing a button with either the index (“Yes”-answer) or the middle (“No”-answer) finger of the right hand. Catch-trials were excluded from all subsequent analyses.

The sounds were played according to a slow-event related design with a variable interstimulus interval (ISI) of 6 to 8 TRs (TR = 2500 ms, average ISI 17.5 s). At the beginning of the fMRI session, the volume of the stimuli was adjusted to a comfortable intensity level. The stimuli were presented in the 500 ms silent gap via MR-compatible earphones (Sensimetrics S14, Malden, MA, USA). After the experiment, all subjects reported a clear hearing of the stimuli. Every run consisted of 5 trials for each of the 9 stimulus conditions and 5 catch-trials, resulting in a total of 50 trials and a run duration of approximately 15 min. The order of stimulus presentation was randomized within and across runs. Four functional runs were acquired, leading to a total of 200 trials in the whole experiment.

2.3.3 Data acquisition

Functional and anatomical images were acquired with a 7T Siemens Magnetom scanner using a 32-channel Nova Medical head coil. Four high-resolution (1.1 mm isotropic voxel size) functional runs were acquired using a gradient-echo (GE) EPI sequence (Moeller et al., 2010) with the following parameters: TR 2500 ms, TE 22 ms, Partial Fourier 5/8, GRAPPA 2, delay in TR 500 ms, multi-band acceleration factor 2 with blipped-CAIPIRINHA (1/FOV shift 4; Setsompop et al., 2012). The sequence was optimized to maximize tSNR in the auditory cortex. In two separate pilot runs of 140 volumes ($\sim$6min, resting-state), we acquired the sequence with these parameters and additionally a variant with GRAPPA 3, Partial Fourier 6/8 and TE 24.4 ms. The latter showed less distortions and signal dropout only in the anterior and posterior parts of the brain albeit with a lower tSNR in the auditory cortex (23.27 versus 34.02, respectively).

In addition to the magnitude images, phase images were collected in order to allow image reconstruction with lower voxel resolution (see below for details). 48 slices were acquired centered approximately on the superior temporal gyrus, covering the auditory cortex. One high-resolution (0.7 mm isotropic voxel size) anatomical image covering the whole brain was collected using MP2RAGE sequence (Marques et al., 2010).

2.3.4 Data analysis: preprocessing and univariate analysis

Functional and anatomical data were preprocessed and analyzed in BrainVoyager QX 2.8.2 (Brain Innovation). The four functional runs were 3D motion corrected and coregistered to the first volume of the first run through rigid-body transformation (3 translational and 3 rotational parameters). Neither nonlinear transformation nor distortion correction algorithm were applied to avoid interpolation confounds in our comparison across resolutions. We visually inspected every coregistered run and no large motion was observed. Linear and low frequency non-linear drifts up to 7 cycles per time course were removed via temporal high-pass filtering. This cut-off frequency, corresponding to a cut-off period of $\sim$128 s, was adequate to the stimulus design and analyses here employed (as estimated through spectral analysis of the class stimulus design). For each subject, the anatomical image was segmented at the gray-white matter boundary via an automatic procedure. The obtained segmentation was visually inspected and manually corrected where necessary. Then, the anatomy was realigned in the subject-specific native space of the functional data in order to avoid any additional resampling and/or interpolation of the functional data. Finally, the segmented anatomical image was inflated to create an individual cortex mesh for each subject. The obtained inflated cortices were used for mask creation and displaying purposes.

Standard univariate statistical analysis of the fMRI data was performed for each subject using general linear modelling of the time series. Voxels with a mean EPI value lower than 100 a.u. were excluded in order to prevent inclusion of cerebrospinal fluid (CSF) voxels in the ROI. The design matrix consisted of a predictor for each vowel and speaker combination convolved with a canonical (double gamma) hemodynamic response function (HRF) in order to account for the hemodynamic response delay. Contrast analysis was performed for each binary comparison of vowels and speakers.

2.3.5 Data analysis: multivariate analysis

Multivariate analysis was performed on a single-subject basis in anatomically defined masks including the superior temporal gyrus (STG), Heschl's gyrus (HG) and superior temporal sulcus (STS). The masks were delineated on the inflated cortex mesh in the native space of each subject.

A linear Support Vector Machine (SVM) (see below) was trained to decode either the vowels (/a/ vs /i/, /a/ vs /u/, /i/ vs /u/) or the speakers (/sp1/ vs /sp2/, /sp1/ vs /sp3/, /sp2/ vs /sp3/). In case of vowel decoding, the trials were grouped according to the vowel irrespectively of the speaker dimension (e.g., the class /a/ contained all the trials corresponding to /sp1_a/, /sp2_a/, and /sp3_a/). In case of speaker decoding, trials were grouped according to the speaker irrespectively of the vowel dimension (e.g., the class /sp1/ contained all the trials corresponding to /sp1_a/, /sp1_i/, and /sp1_u/). Three classes of 60 trials were obtained in both cases. For each comparison, the classification was performed in leave-one-run-out cross-validation. At each fold of the cross-validation 3 runs were used for feature selection, feature extraction and training of the classifier and 1 run to test the performance. The details of each of these stages are explained in the following paragraphs.

Feature selection

For the purpose of algorithm efficiency and interpretability of the results, we further restricted the selection of voxels entering the classification by univariately choosing the n most active voxels per class. As MVPA classification accuracy may depend on number of selected voxels (Cox and Savoy, 2003), we let n vary between 50 and 10000 voxels in discrete steps (n = [50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 3000, 4000, 5000, 6000, 8000, 10000]) for a total of 19 levels of feature selection. The n most active voxels per class were defined as the n voxels with highest t-values resulting from the GLM of the training set. Note that the training set of a specific binary comparison and a specific cross-validation fold contained the trials corresponding to the two classes of that comparison and the 3 training runs of that fold. Considering for example the binary comparison /sp1/ vs /sp2/ and the 1^{st} cross-validation fold (in which run #1 is used for testing), all the trials corresponding to the stimuli /sp1_a/, /sp1_i/ and /sp1_u/ from the run #2, #3 and #4 were included in the predictor of /sp1/ and all the trials corresponding to the stimuli /sp2_a/, /sp2_i/ and /sp2_u/ from the run #2, #3 and #4 were included in the predictor of /sp2/. Once the GLM was performed, we selected the voxels having the n highest t-value for the predictor /sp1/ and those having the n highest t-value for the predictor /sp2/. The union of these two ensemble of voxels, which was ranging between n and $2 * n$ depending on the amount of overlap between the two ensembles, represented the voxels selected

for that specific level of feature selection n and was passed to the *feature extraction* stage.

Feature extraction

The preprocessed time courses of the selected voxels were divided into single-trials according to the time of presentation of each stimulus. Each trial was normalized to the averaged value of the signal at 2500 ms (1 TR) and 0 ms before the stimulus onset to avoid biases due to global changes in the baseline signal across runs or signal drifts within and across runs. The obtained percent BOLD signal change was fitted with an optimized model of hemodynamic response function resulting in a β-value for each trial and each voxels which was then used as a feature in the classifier.

The HRF model used here was optimized to take into account potential differences in the temporal delay of the BOLD response. The optimal delay for each voxel was calculated by averaging the hemodynamic response across all trials and fitting it with a canonical HRF model whose time-to-peak parameter was varied between 4.0 s and 6.0 s in step of 0.5 s: the time-to-peak giving the best fit (as estimated by F-statistics of the regression) was chosen as optimal HRF delay of that voxel.

Classification

We used the linear SVM implemented in the Spider toolbox (`http://people.kyb.tuebingen.mpg.de/spider/`) as supervised classification algorithm with a fixed regularization parameter (C = 1; see Cortes and Vapnik, 1995, for details of the algorithm). Training and testing of the algorithm were performed in leave-one-run-out cross-validation: 3 runs (i.e., 45 trials per class) were used to train the classifier and 1 run (i.e., 15 trials per class) to test the performance. Voxel selection, feature extraction, training and testing were repeated 4 times for all possible training and testing run combinations. The decoding performance (generalization score) was assessed by averaging the classification accuracies across the 4 folds of the cross-validation.

Finally, the entire classification procedure (including voxel selection, feature extraction, training and testing of the classification in cross-validation) was repeated for each of the nineteen levels of feature selection.

Statistical testing

The statistical significance of the decoding accuracies was assessed at group level using non-parametric statistics. For each subject, we estimated the mean accuracy under the null hypothesis by performing a permutation test with 200 permutations. Note that this limited number of permutations does not allow reliable testing for significance at the single-subject level; however, it provides a stable estimate of the mean of the null distribution, which is then used in the second level analysis. For each permutation, the labels of the training dataset were scrambled using the $randperm$ function of MATLAB (MathWorks, Natick, MA, USA) and the classification procedure repeated at the different feature selection levels. Using the same labels at different feature selection levels allows correctly taking into account the correlations of decoding performance among different levels of feature selection.

In the second level analyses, the accuracy curve (i.e., the set of accuracy values at all nineteen levels of feature selection) was considered in a permutation test. In each permutation, we randomly swapped the accuracy curve of one (or more) subject(s) between the tested model and its null distribution mean and we counted the number of occurrences where the distance between the newly obtained group average curves was larger than the original one. The test was performed with an exact number of permutations $nr_perm = 2^N = 1024$ (with N standing for number of subjects), which allowed to assess a minimum p-value of 2^{-10}. For details of the permutation test at group level, please, see the Supplementary material 2.7.

2.3.6 Effect of spatial smoothing

In order to assess the effect of spatial smoothing of fMRI data on the decoding performance, we spatially smoothed the data with a three-dimensional Gaussian kernel of full-width-half-maximum (FWHM) of 1.1 mm, 2.2 mm, 3.3 mm, 4.4 mm and 8.8 mm. We performed the multivariate analysis on the unsmoothed data and after smoothing with each kernel size.

2.3.7 Reconstruction of data at lower effective spatial resolution

The raw data of the EPI images are acquired in the k-space where each k-value represents a sinusoidal spatial frequency. The inverse discrete Fourier transform

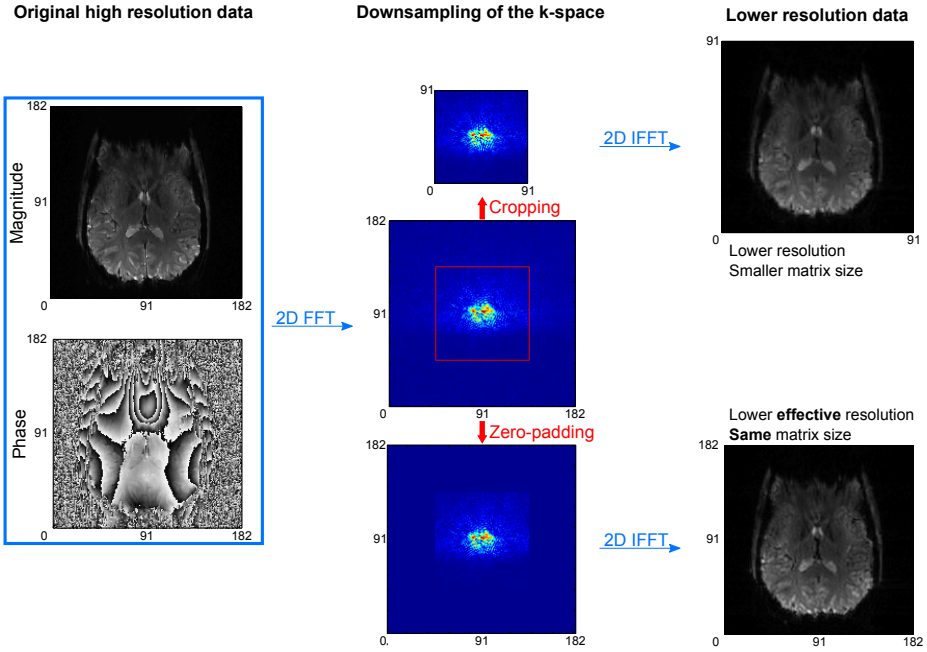

FIGURE 2.1: Pipeline of the reconstruction of the fMRI data at different spatial resolutions (see text for details).

(DFT) of the k-space data yields a complex image with both real and imagery components. These complex components are generally converted into a magnitude and phase image. Typical fMRI studies analyze only the magnitude image, while the phase image is usually discarded (but see Balla et al., 2014, and references therein). The original k-space data can be recovered by applying DFT on the combined information of the magnitude and phase data. The spatial resolution in the image space corresponds to the field-of-view (FOV) in the k-space and vice versa.

In order to determine the influence of spatial resolution on the decoding performance, we reconstructed the data at different lower effective resolutions, as follows (see Figure 2.1): First, for each slice, we used both the magnitude and phase images collected at the original high resolution (1.1 mm isotropic) to reconstruct the complex k-space via 2-dimensional fast-Fourier transform (FFT). Second, we retained a central submatrix of the k-space and we substituted the rest of the k-values with zero. In this way, we effectively removed the high spatial frequency information while leaving the low spatial frequency information

unaffected. Finally, we applied the inverse 2D FFT to obtain an image with the same nominal resolution albeit with high spatial frequency information removed, i.e., lower effective resolution.

The original matrix size was 182x182 and the reconstructed ones had non-zero k-space information in the central 91x91 and 61x61 submatrices leading to an effective in-plane resolution of 2.2x2.2 mm^2 and 3.3x3.3 mm^2, respectively. This approach mimics the acquisition of fMRI data at different in-plane resolutions since the downsampling of the data is performed in the complex k-space. However, zeroing the external part of the k-space matrix, instead of cropping it, allowed us to preserve the same matrix size and, therefore, the same number of voxels in all analyses.

Note that the downsampling approach here presented is different from smoothing techniques in the image space (such as Gaussian smoothing), although both are filtering techniques. Gaussian smoothing is an invertible operation, which attenuates high spatial frequency components, but it does not eliminate them entirely (Kamitani and Sawahata, 2010; Swisher et al., 2010). Our downsampling methodology is a non-invertible operation, which directly manipulates the information in the spatial frequency space by complete cut-off of the high frequency components. As a control analysis, the effect of resolution was tested also using downsampling of the complex data through cropping (see supplementary material for details and results). Importantly, not only the smoothing kernel is different for both approaches but also the data on which these operations are applied (magnitude vs. magnitude and phase data). Fourier transforming the magnitude data would not yield the complex k-space sampled during fMRI acquisition and would, therefore, not allow mimicking acquisition at lower (effective) spatial resolutions via zeroing/cropping its external part.

We preprocessed (motion correction, high-pass filtering and spatial smoothing) and analyzed the datasets at lower effective resolutions following the same procedures as described above for the original data. The effects of resolution and smoothing on decoding accuracies were tested via a two-way repeated measures analysis of variance using R (package ez, `http://CRAN.R-project.org/package=ez`).

2.3.8 Influence of tSNR, CNR and head motion on MVPA

Standard machine learning techniques rely on estimating features from a training set in order to predict labels in an unseen testing set. In fMRI MVPA, this translates in the assumption that the experimental stimuli drive the same pattern of activation through the entire duration of the experiment from which training and testing dataset are extracted. However, stimulus related signal is modulated by instrumental noise and subject-related noise, such as physiological fluctuations and motion artifacts. In order to assess the influence of these factors on MVPA performances, we investigated the relationship between decoding accuracies and a) temporal signal-to-noise ratio (tSNR), a measure of both instrumental and physiological noise; b) functional contrast-to-noise ratio (CNR), a measure of the strength of the activation signal with respect to noise and c) head motion.

TSNR of the fMRI signal in the auditory cortex was calculated as the ratio between the mean fMRI signal and its temporal standard deviation and averaged across all voxels in the mask. CNR was calculated as the ratio between the standard deviation of the fMRI signal response and the standard deviation of the baseline (Welvaert and Rosseel, 2013). CNR was calculated only on the active voxels as defined in the feature selection step of the decoding procedure. Head motion was quantified on the basis of the rotation and translation parameters obtained from the motion correction algorithm: each motion parameter was linearly detrended; then, its standard deviation was calculated; finally, the average of the standard deviations of all parameters was taken as motion index.

2.3.9 Voxel-wise tissue type ranking

To characterize the physiological origin of multivariate information in fMRI, voxels were ranked according to their likelihood of containing gray matter (i.e., microvasculature) or large blood vessels (i.e., macro-vasculature). Such likelihood was calculated following the method introduced by Shmuel et al. (2010). First, voxels were ranked separately according to two different quantities: their EPI value and their BOLD percent signal change. Both lower EPI values (with no apparent field distortions and susceptibility artifacts in the ROI) and higher BOLD percent changes indicate large veins (see Uludag et al., 2009). Then, a global rank was built by averaging the "BOLD percent change rank" and the inverse of the "EPI value rank". Voxels were, thus, sorted according to their most likely tissue type:

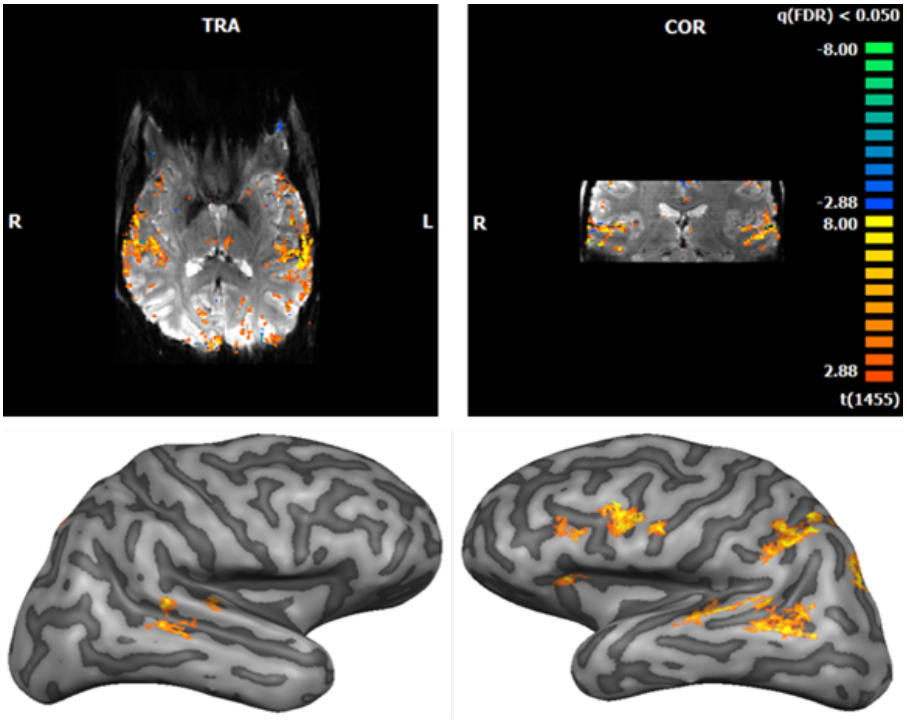

FIGURE 2.2: BOLD activation in response to speech sounds as resulting from GLM analysis of unsmoothed data (single-subject, statistical parametric F-map, Q[FDR] <0.05 and cluster size threshold of 4 voxels). Top row, transversal (TRA) and coronal (COR) view; bottom row, projection of the F-map on the inflated cortex (light and dark gray represents gyri and sulci, respectively).

from gray matter (low global rank) to large vessels (high global rank) regions. For the rest of the paper, we will call this global rank *blood vessels* (BV) *likelihood*.

2.4 Results

2.4.1 Univariate analysis

Univariate analysis showed significant BOLD signal activation (Q[FDR] <0.05 and cluster size threshold of 4 voxels; FDR stands for false discovery rate) in the auditory cortex in response to the stimulus sounds (Figure 2.2), similar to (Formisano et al., 2008). Outside the temporal lobe, significant activation was observed in the frontal (Broca's area) and parietal lobe. Pair-wise contrast analysis

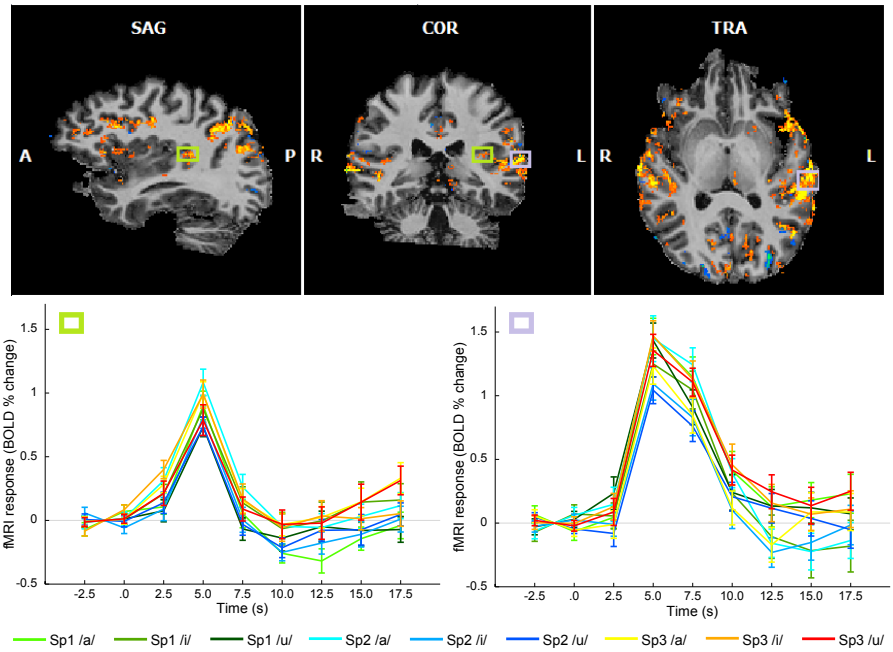

FIGURE 2.3: Examples of mean BOLD response for different areas of the auditory cortex: left Heschl's gyrus (green ROI) and superior temporal sulcus (violet ROI). Curves of different colors represent the mean BOLD response activated by the 9 different stimuli.

between speakers and/or vowels did not show significant effects for any comparison (data not shown).

Locally averaged BOLD responses to the sound stimuli exhibit large variability in the shape and latency of the hemodynamic response in different areas of the auditory cortex. In general, the fMRI responses from the Heschl's gyrus and proximities have narrower response width than the responses from the auditory belt areas (see Figure 2.3).

2.4.2 Vowel and speaker decoding from the original data

Multivariate analysis was performed on the original fMRI data (1.1 mm isotropic spatial resolution) after preprocessing with 2.2 mm Gaussian smoothing. In the Section 2.4.3 *Effects of spatial resolution and smoothing on decoding*, we will motivate this choice demonstrating that smoothing using a kernel size of 2.2 mm yields optimal preprocessing of the data.

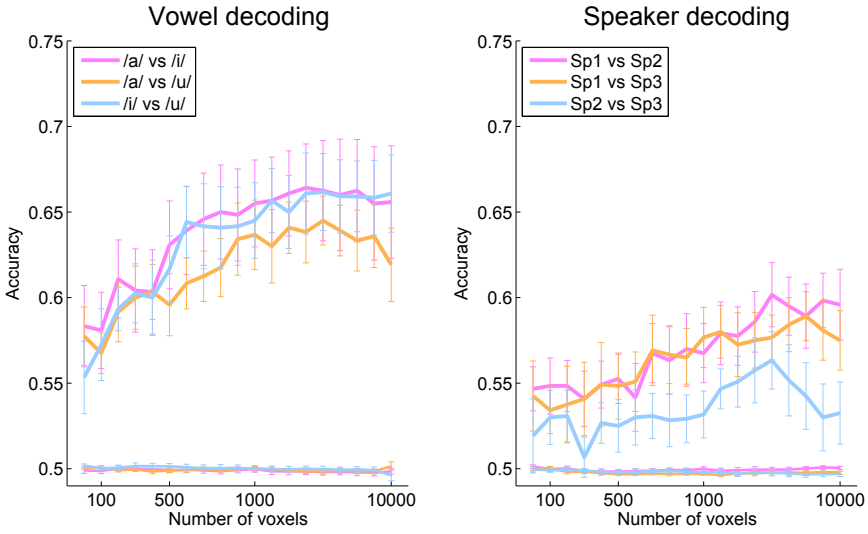

FIGURE 2.4: Group level decoding accuracy for the data at the original resolution of 1.1 mm isotropic after smoothing with 2.2 mm Gaussian kernel. Accuracies are plotted as a function of the number of voxels selected. Each line represents the mean accuracy averaged across all the subjects and error bars represent the standard error of the mean. Thinner lines around 0.5 represent the corresponding results for the decoding with scrambled labels.

Figure 2.4 shows the group level results of vowel and speaker decoding (left and right, respectively) performed on the data with resolution of 1.1 mm isotropic. For each binary comparison, we plotted the average accuracy as a function of the number of voxels selected. The accuracy curves of all models show a similar dependency on the number of voxels selected: accuracy first increases with the number of voxels and then reaches a plateau. In some cases, accuracy decreases for extremely large number of voxels.

Every model was decoded above chance level as verified with permutation tests at group level (chance level is represented in the plots with thinner lines). Vowels could be robustly decoded reaching an average accuracy of 64%, 62%, and 63% for /a/ vs /i/, /a/ vs /u/ and /i/ vs /u/, respectively ($p < 0.001$ for all comparisons, permutation test). Speakers were decoded with a lower but significant average accuracy of 57% for Sp1 vs Sp2 ($p = 0.0029$), 56% for Sp1 vs Sp3 ($p < 0.001$) and 53% for Sp2 vs Sp3 ($p = 0.0049$).

All vowels were decoded with comparable performances, while decoding between the female speaker (Sp1) versus one male speaker (Sp2 and Sp3) resulted in significantly higher accuracy than decoding between the two male speakers: 'Sp1 vs Sp2' >'Sp2 vs Sp3' with p-value = 0.0137, 'Sp1 vs Sp3' >'Sp2 vs Sp3' with p-value = 0.0059 (as assessed via permutation test).

2.4.3 Effects of spatial resolution and smoothing on decoding

For each combination of spatial resolution and Gaussian smoothing preprocessing, we repeated the entire multivariate procedure including all levels of feature selection. For display purposes, we averaged the group accuracies across the levels of feature selection and present them in Figure 2.5 by a surface as a function of spatial resolution and smoothing kernel size.

The effects of resolution and smoothing on decoding accuracies were tested via a two-way repeated measures analysis of variance. In the vowel decoding case, both main effects and their interaction were significant (main effect of resolution: $F(2, 18) = 10.14$, $p = 0.001$; main effect of smoothing: $F(5, 45) = 13.65$, $p < 0.001$; interaction effect between resolution and smoothing: $F(10, 90) = 2.67$, $p = 0.007$). In the speaker decoding case, only the main effect of smoothing was significant (main effect of resolution: $F(2, 18) = 1.58$, $p = 0.233$; main effect of smoothing: $F(5, 45) = 4.45$, $p = 0.002$; interaction effect between resolution and smoothing: $F(10, 90) = 1.02$, $p = 0.433$). Control analyses (see Supplementary material 2.6) were performed to assure that these results were not biased by considering the average accuracy across the levels of features selection, as they do not depend on the specific level of feature selection considered.

For both vowel and speaker decoding, the spatial smoothing improved the decoding accuracies in case of moderate smoothing (up to 2.2 mm), while more smoothing decreased them. This suggests that the optimal preprocessing of the fMRI data for decoding analysis, at least for the present data set, is obtained with a spatial smoothing of 2.2 mm FWHM 3D Gaussian kernel. For this reason, analyses in the rest of the paper were performed on data preprocessed with 2.2 mm FWHM Gaussian smoothing, unless otherwise specified.

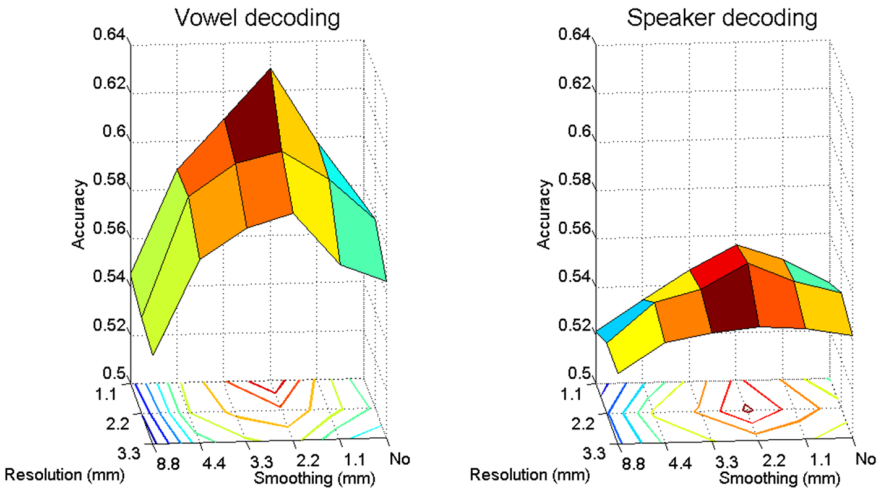

FIGURE 2.5: Resolution and smoothing effect on decoding performances. Each point of the surface reports the group level accuracy of vowel (left plot) and speaker (right plot) decoding as a function of the spatial resolution of the data (from back to front: 1.1x1.1, 2.2x2.2 and 3.3x3.3 mm^2 in-plane) and the kernel size of spatial smoothing applied them (from right to left: unsmoothed and smoothed with Gaussian kernel size 1.1, 2.2, 3.3, 4.4, 8.8 mm). For display purposes, the accuracy is shown as the result of averaging across all levels of feature selection.

2.4.4 Effect of spatial resolution on decoding

Figure 2.6 illustrates in more detail the impact of acquiring fMRI data at different spatial resolutions on decoding performances. Each curve of the plots represents the group level decoding accuracy for each spatial resolution, calculated as the average across the three models of vowel or speaker decoding (left and right panel, respectively) and subjects. The stimuli were decoded significantly above chance level ($p < 0.01$, permutation test) at every resolution for both vowels and speakers. However, the decoding performances for vowels and speakers showed a different dependency on the effective spatial resolution of the analyzed data. Vowel decoding accuracies significantly increased with the increase in effective spatial resolution (OR1>ER2 $p = 0.0078$, OR1>ER3 $p = 0.0020$, ER2>ER3 $p = 0.0498$; permutation test; OR1 stands for original resolution 1.1 mm, ER2 for effective resolution 2.2 mm and ER3 for effective resolution 3.3 mm). In contrast, speaker decoding showed no significant accuracy differences between the tested resolutions, although a tendency of higher accuracies was observed for

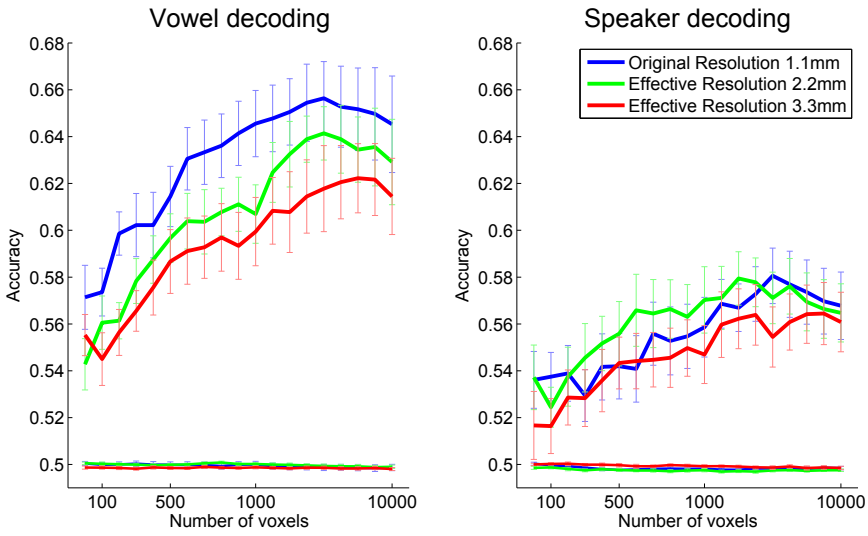

FIGURE 2.6: Resolution effect on decoding accuracy. Each line represents the decoding accuracy averaged across all models of vowel or speaker decoding and all subjects for a certain resolution. Thinner lines around 0.5 represent the corresponding results for the decoding with scrambled labels. Error bars represent the standard error of the mean across the subjects. Both vowels and speakers are decoded above chance for all resolutions. In the left panel, vowel decoding performance significantly increases with increasing resolution: OR1>ER2 $p = 0.0078$, OR1>ER3 $p = 0.0020$, ER2>ER3 $p = 0.0498$ (OR1 stands for original resolution 1.1 mm, ER2 for effective resolution 2.2 mm and ER3 for effective resolution 3.3 mm). In the right panel, speaker decoding performance was almost significantly better for the effective resolution of 2.2 mm with respect to the effective resolution of 3.3 mm: OR1>ER2 $p = 0.6777$, OR1>ER3 $p = 0.1748$, ER2>ER3 $p = 0.0508$. p-values were calculated via permutation test as explained in the Section 2.3.5 *Statistical testing* with the only difference that in this case we randomly swapped accuracy curves between the tested resolutions (instead of between "real" accuracy curve and "chance" accuracy curve).

the effective resolution of 2.2 mm with respect of the 3.3 mm dataset (ER2>ER3 $p = 0.0508$; permutation test).

The different resolution dependency for vowel and speaker decoding was confirmed by a permutation test ($nr_perm = 2^{10}$) on the differences between accuracy curves at the investigated resolutions. Note that considering the difference between accuracy curves allows to account for a potential confound due to global performance differences in vowel and speaker decoding. The resolution

effect on vowel decoding was significantly larger than that on speaker decoding for the difference between original and 2.2 mm in-plane resolution ($p = 0.0088$), close to significance for the difference between 1.1 and 3.3 mm in-plane resolution ($p = 0.0762$), and not significant for the difference between 2.2 and 3.3 mm in-plane resolution ($p > 0.1$).

Accuracy curves for the two lower effective resolutions showed a dependency on the number of voxels selected similar to the one of the original resolution (described in the Section 2.4.2 *Vowel and speaker decoding from the original data* and depicted in Figure 2.6 by the blue line).

2.4.5 Influence of tSNR, CNR, and head motion on MVPA

We studied the relationship between averaged accuracies and temporal signal-to-noise ratio (tSNR), functional contrast-to-noise ratio (CNR) and head motion. Although no correlation was significant, we found that in the case of vowel decoding (and similarly for the speaker decoding case) the averaged accuracy tends to be positively correlated with tSNR ($r = 0.55$, $p = 0.09$) and CNR ($r = 0.37$, $p > 0.1$) while negatively correlated with head movement ($r = -0.34$, $p > 0.1$). Note that both increasing spatial smoothing and lowering spatial resolution of the fMRI data resulted in a monotonous increase of tSNR and CNR (data not shown).

2.4.6 Influence of tissue type on decoding

Figure 2.7 shows the weights given by the classifier to each voxel as a function of BV likelihood (from 0% indicating most likely gray matter voxels to 100% indicating most likely large blood vessel voxels) for the case of vowel decoding. Weights are on average slightly, but significantly, higher for the 10% "more GM" voxels than for the 10% "more large BV" voxels (0.040 ± 0.006 versus 0.034 ± 0.007, respectively; t-test, p-value $= 0.0044$). Similar results are obtained in case of speaker decoding (0.036 ± 0.005 versus 0.031 ± 0.007, respectively; t-test, p-value $= 0.0035$).

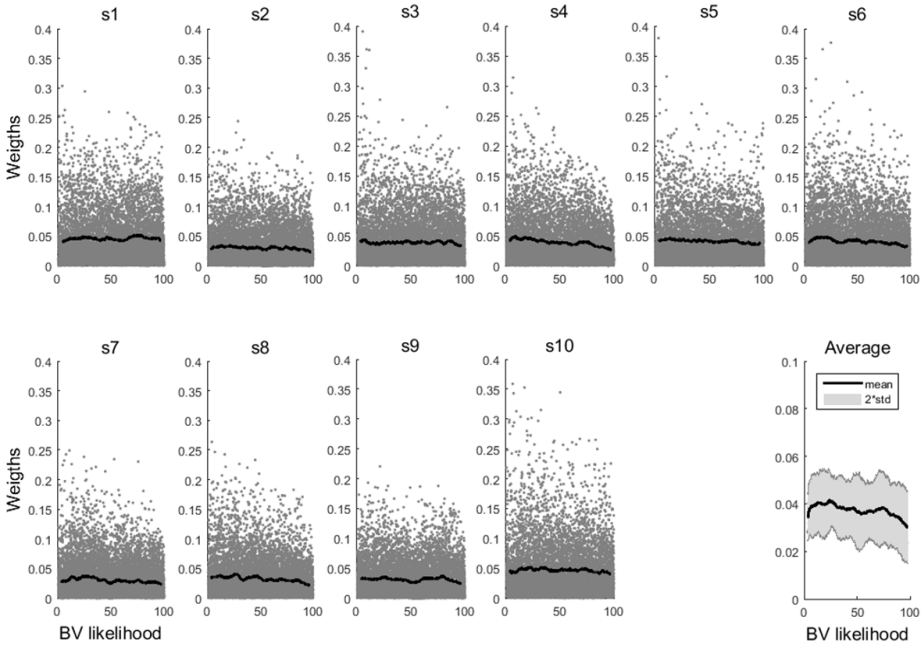

FIGURE 2.7: Distribution of classification weights versus BV likelihood for each subject in the case of vowel decoding. Each gray dot represents one voxel and the black line the moving average of the weights calculated excluding the 5% more extreme values in order to avoid boundary effects. The bottom right plot shows the group moving average (black line) and its standard deviation (light gray shadow).

2.5 Discussion

In this study, we used high-resolution 7T fMRI data and their reconstruction at different effective spatial resolutions to investigate the relevant spatial scale of information content of the fMRI signal in the auditory cortex for vowel and speaker identity. To the best of our knowledge, this work is the first investigating the relation between spatial resolution and MVPA with fMRI at ultra-high magnetic field strength. Instead of acquiring additional data at lower spatial resolutions, we simulated their acquisition starting from the high-resolution dataset. This allowed us to compare decoding performance for the very same dataset as a function of effective spatial resolution. In this way, we could exclude variability across runs and sessions related to subject performance (e.g., attention to the stimuli, movement) and acquisition differences (e.g., coil sensitivity, scanner drift).

We performed multivariate analysis of the original high-resolution data (1.1 mm isotropic) and demonstrated the feasibility of decoding speech content and speaker identity from 7T fMRI data in the auditory cortex. These results are in agreement with the findings by (Formisano et al., 2008), which demonstrated at 3T the feasibility to decode vowels and speakers from auditory cortical response patterns. Note that a direct comparison of accuracies as obtained in this study with those in (Formisano et al., 2008) is difficult due to the different methods of cross-validation and feature selection employed. Formisano and colleagues used a leave-k-stimulus-pair-out cross-validation (see Formisano et al., 2008, for details), while we employed a leave-one-run-out cross-validation. This latter approach is more conservative because it controls for false-positive rate at the nominal level; thus, it may result in lower accuracy values (Misaki et al., 2010). Moreover, Formisano and colleagues used a combination of univariate and multivariate feature selection strategies to improve sensitivity (see De Martino et al., 2008; Formisano et al., 2008, for details). Here, we chose a univariate only feature selection method because of computational complexity, which is of particular concern for high-resolution data and large number of features (voxels).

To overcome an arbitrary choice of number of features selected, the multivariate analysis was repeated for several feature selection levels (from 50 to 10000 voxels) and decoding accuracies of different models were compared by considering the resulting decoding accuracy curves. The accuracy curves consistently showed a dependency on the number of voxels selected: accuracy increased with the number of voxels, reached a plateau, and generally decreased for extremely high number of voxels. This observation is in agreement with previous studies (Cox and Savoy, 2003; Fan et al., 2007; De Martino et al., 2008) and highlights that our results do not critically depend on a specific feature selection level. In general, the optimal number of voxels to be included in the MVPA analysis may depend on the spatial resolution, the decoding task and image characteristics. Additionally, we observed that speaker decoding was performed with higher accuracy for comparison between the female and a male speaker than between the two male speakers. These results reflect the smaller difference in fundamental frequency among the two male speakers (compared to the female speaker, see also Formisano et al., 2008, cfr. Figure 2b).

The comparison of MVPA performances at different spatial resolutions showed that decoding accuracies increased with increasing resolution in case of vowel decoding. In case of speaker decoding, however, such behavior was not observed. Instead, there was a tendency to observe the highest accuracy for the intermediate spatial resolution. These results show that high spatial frequencies are informative for vowel decoding and that the relative contribution of high and low spatial frequencies is different across the two decoding tasks. In turn, this suggests a distinct spatial arrangement of the distributed neural/neuro-vascular sources underlying vowel and speaker representation and processing in auditory cortex.

These outcomes are consistent with previous observations that informative voxels in the auditory cortex are widely distributed for vowel decoding, while more clustered for speaker decoding (Formisano et al., 2008; Bonte et al., 2014). In addition, they support the hypothesis that the brain response patterns enabling the decoding of vowels and speakers reflect the representation and processing of distinct type of acoustic information (e.g., the representation of "formants combination" for vowel decoding and of "fundamental frequency" for speaker decoding, see Formisano et al., 2008). Furthermore, the finding that the maximum information (MVPA accuracy) is decoded optimally at different resolutions for vowels than for speakers provides a clear example of the complex dependency of MVPA results on the underlying physiological basis and thus of the need to address the issue of optimal acquisition strategy case-by-case through empirical studies (Formisano and Kriegeskorte, 2012).

In contrast to the resolution effect, we noted that the effect of smoothing is similar for vowel and speaker decoding and for different resolutions. This result suggests that the optimal smoothing kernel depends mainly on intrinsic noise properties of the data. This is in agreement with previous findings by Kamitani and Sawahata (2010). They showed that smoothing (e.g., with a Gaussian kernel) and more generally spatial convolution does not remove information content as it is an invertible transformation. However, such equivalence between original and smoothed data in terms of information content, and therefore classification results, holds only in ideal cases where the signal from neighboring voxels is entirely independent, no selection of voxels is made and no motion (or other sources of) artifacts are present. Thus, in experimental fMRI data, the influence of smoothing on decoding might be attributed to the attenuation of (thermal)

noise due to the signal correlation among neighboring voxels. Moreover, we observed that tSNR and CNR monotonously increased with larger smoothing kernels, while the decoding accuracies peaked at 2.2 mm 3D Gaussian kernel size. Therefore, the smoothing effect cannot be explained only on the basis of these two factors. Interestingly, such optimal spatial smoothing was reached for a kernel size of 2.2 mm which is similar to the experimentally-determined point spread function at 7T (Shmuel et al., 2007). Another reason why moderate smoothing is beneficial for MVPA could be the residual head motion present in the data even after motion correction. This hypothesis is consistent with previous studies showing that head motion reduces MVPA performances (Kriegeskorte et al., 2010; Swisher et al., 2010; Alink et al., 2013; Misaki et al., 2013). On the other hand, smoothing with large kernel sizes results in poorer MVPA performances due to partial volume effects with white matter and CSF or gray matter on an opposing sulcal bank (Andrade et al., 2001). For this reason, smoothing along the 2D cortical sheet may be more beneficial than performing spatial smoothing in the volume. Our results showing that large smoothing kernels are detrimental for MVPA decoding are in line with previous studies (Kriegeskorte et al., 2010; Swisher et al., 2010; Alink et al., 2013; Misaki et al., 2013). However, Op de Beeck (2010) reported no influence of spatial smoothing on decoding performances. A possible explanation of these different outcomes may be the different preprocessing of the functional data. In Op de Beeck (2010), data were coregistered in the anatomical space of the subject and then normalized to MNI template (Montreal Neurological Institute). These two transformations are achieved via re-sampling and interpolation of the original functional data. Therefore, the smoothness introduced in the transformed data could make the manipulation through different spatial smoothing kernels less sensitive, especially in experiments where decoding accuracies are at ceiling level (100% decoding accuracy for the object decoding experiment).

One of the advantages of using a linear SVM algorithm is the possibility to interpret the weights attributed to each voxel in terms of discriminability power. In order to shed further light on the physiological origin of the multivariate information in fMRI, we ranked the voxels according to their tissue properties. We showed that discriminability power is distributed over the continuum of tissue type ranging from gray matter to large blood vessels regions, with a slightly higher contribution from gray matter than from blood vessel voxels. This result suggests that decodable information resides both in the tissue and large

veins. Future studies, possibly including venograms and/or using different type of functional contrast such as spin echo, arterial spin labeling, or vascular space occupancy, are needed to understand the role of tissue types in fMRI MVPA.

Another factor that could influence MVPA performance in fMRI, is HRF variability. In this study, we observed important changes in the HRF shape in different regions. However, a systematic investigation of this phenomenon and its influence on MVPA is beyond the scope of this paper and needs to be addressed in future developments.

In summary, in this study, we investigated MVPA in fMRI at 7T. Our results suggest that different features of complex (auditory) stimuli are spatially distributed in the brain and may be represented at different spatial scales. As a general consequence, there is no optimal spatial resolution for all decoding tasks but the acquisition needs to be tailored to the stimuli and brain area investigated. However, our findings for vowel decoding are compatible with the hypothesis that MVPA information partly originates from spatial arrangements of neuronal populations at a spatial scale below fMRI resolution (< 1 mm). This spatial information may become available at resolutions accessible for fMRI due to local irregularities and/or transposition by hemodynamic filtering (Chaimow et al., 2011). Under this hypothesis and in the absence of precise knowledge of the spatial scale of interest, we recommend acquisition of data at high spatial resolution (1 mm or below). This enables the detection of stimulus-related information at all measurable spatial frequencies, which is required for inferring the spatial organization of the underlying neuronal architecture.

Bibliography

A. Alink, A. Krugliak, A. Walther, and N. Kriegeskorte. fMRI orientation decoding in V1 does not require global maps or globally coherent orientation stimuli. *Front Psychol*, 4:493, 2013.

A. Andrade, F. Kherif, J. F. Mangin, K. J. Worsley, A. L. Paradis, O. Simon, S. Dehaene, D. Le Bihan, and J. B. Poline. Detection of fMRI activation using cortical surface mapping. *Hum Brain Mapp*, 12(2):79–93, 2001.

D. Z. Balla, R. M. Sanchez-Panchuelo, S. J. Wharton, G. E. Hagberg, K. Scheffler, S. T. Francis, and R. Bowtell. Functional quantitative susceptibility mapping (fQSM). *Neuroimage*, 100:112–24, 2014.

M. Bonte, L. Hausfeld, W. Scharke, G. Valente, and E. Formisano. Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *J Neurosci*, 34(13):4548–57, 2014.

G. M. Boynton. Imaging orientation selectivity: decoding conscious perception in V1. *Nat Neurosci*, 8(5):541–2, 2005.

D. Chaimow, E. Yacoub, K. Ugurbil, and A. Shmuel. Modeling and analysis of mechanisms underlying fMRI-based decoding of information conveyed in cortical columns. *Neuroimage*, 56(2):627–42, 2011.

C. Cortes and V. Vapnik. Support-Vector Networks. *Machine Learning*, 20(3):273–297, 1995.

D. D. Cox and R. L. Savoy. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*, 19(2 Pt 1):261–70, 2003.

F. De Martino, G. Valente, N. Staeren, J. Ashburner, R. Goebel, and E. Formisano. Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage*, 43(1):44–58, 2008.

Y. Fan, H. Rao, H. Hurt, J. Giannetta, M. Korczykowski, D. Shera, B. B. Avants, J. C. Gee, J. Wang, and D. Shen. Multivariate examination of brain abnormality using both structural and functional MRI. *Neuroimage*, 36(4):1189–99, 2007.

E. Formisano and N. Kriegeskorte. Seeing patterns through the hemodynamic veil - the future of pattern-information fMRI. *Neuroimage*, 62(2):1249–56, 2012.

E. Formisano, F. De Martino, M. Bonte, and R. Goebel. "Who" is saying "what"? brain-based decoding of human voice and speech. *Science*, 322(5903):970–3, 2008.

J. Freeman, G. J. Brouwer, D. J. Heeger, and E. P. Merriam. Orientation decoding depends on maps, not columns. *J Neurosci*, 31(13):4792–804, 2011.

J. L. Gardner. Is cortical vasculature functionally organized? *Neuroimage*, 49(3): 1953–6, 2010.

J. L. Gardner, P. Sun, D. Tanaka, D. J. Heeger, and K. Cheng. Classification analysis with high spatial resolution fMRI reveals large draining veins with orientation specific responses. *Society for Neuroscience Meeting, Atlanta, GA, USA*, 2006.

J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–30, 2001.

J. D. Haynes and G. Rees. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci*, 8(5):686–91, 2005.

J. D. Haynes and G. Rees. Decoding mental states from brain activity in humans. *Nat Rev Neurosci*, 7(7):523–34, 2006.

Y. Kamitani and Y. Sawahata. Spatial smoothing hurts localization but not information: pitfalls for brain mappers. *Neuroimage*, 49(3):1949–52, 2010.

Y. Kamitani and F. Tong. Decoding the visual and subjective contents of the human brain. *Nat Neurosci*, 8(5):679–85, 2005.

Y. Kamitani and F. Tong. Decoding seen and attended motion directions from activity in the human visual cortex. *Curr Biol*, 16(11):1096–102, 2006.

N. Kriegeskorte and P. Bandettini. Analyzing for information, not activation, to exploit high-resolution fMRI. *Neuroimage*, 38(4):649–62, 2007.

N. Kriegeskorte, R. Cusack, and P. Bandettini. How does an fMRI voxel sample the neuronal activity pattern: compact-kernel or complex spatiotemporal filter? *Neuroimage*, 49(3):1965–76, 2010.

J. P. Marques, T. Kober, G. Krueger, W. van der Zwaag, P. F. Van de Moortele, and R. Gruetter. MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *Neuroimage*, 49(2):1271–81, 2010.

M. Misaki, Y. Kim, P. A. Bandettini, and N. Kriegeskorte. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage*, 53(1):103–18, 2010.

M. Misaki, W. M. Luh, and P. A. Bandettini. The effect of spatial smoothing on fMRI decoding of columnar-level organization with linear support vector machine. *J Neurosci Methods*, 212(2):355–61, 2013.

S. Moeller, E. Yacoub, C. A. Olman, E. Auerbach, J. Strupp, N. Harel, and K. Ugurbil. Multiband multislice GE-EPI at 7 Tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn Reson Med*, 63(5):1144–53, 2010.

H. P. Op de Beeck. Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? *Neuroimage*, 49(3):1943–8, 2010.

J. R. Polimeni, B. Fischl, D. N. Greve, and L. L. Wald. Laminar analysis of 7T BOLD using an imposed spatial activation pattern in human V1. *Neuroimage*, 52(4):1334–46, 2010.

K. Setsompop, B. A. Gagoski, J. R. Polimeni, T. Witzel, V. J. Wedeen, and L. L. Wald. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magn Reson Med*, 67 (5):1210–24, 2012.

A. Shmuel, E. Yacoub, D. Chaimow, N. K. Logothetis, and K. Ugurbil. Spatio-temporal point-spread function of fMRI signal in human gray matter at 7 Tesla. *Neuroimage*, 35(2):539–52, 2007.

A. Shmuel, D. Chaimow, G. Raddatz, K. Ugurbil, and E. Yacoub. Mechanisms underlying decoding at 7 T: ocular dominance columns, broad structures, and macroscopic blood vessels in V1 convey information on the stimulated eye. *Neuroimage*, 49(3):1957–64, 2010.

J. D. Swisher, J. C. Gatenby, J. C. Gore, B. A. Wolfe, C. H. Moon, S. G. Kim, and F. Tong. Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. *J Neurosci*, 30(1):325–30, 2010.

K. Uludag, B. Muller-Bierl, and K. Ugurbil. An integrative model for neuronal activity-induced signal changes for gradient and spin echo functional imaging. *Neuroimage*, 48(1):150–65, 2009.

M. Welvaert and Y. Rosseel. On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data. *PLoS One*, 8(11):e77089, 2013.

E. Yacoub, P. F. Van De Moortele, A. Shmuel, and K. Ugurbil. Signal and noise characteristics of Hahn SE and GE BOLD fMRI at 7 T in humans. *Neuroimage*, 24(3):738–50, 2005.

2.6 Supplementary material 1

2.6.1 Changing resolution by cropping the k-space and its influence on decoding

In this study, we reconstructed the original high-resolution (1.1 mm isotropic) data at lower resolutions in order to investigate the influence of spatial resolution on decoding performances. Here, we therefore compared MVPA performances across different resolution also in the case of lower nominal resolutions (i.e., characterized by smaller matrix sizes and total number of voxels). Downsampling of the original high-resolution data at lower nominal resolutions was achieved with the same procedure described in the paper (see Figure 2.1), but cropping the external part of the k-space instead of zeroing it. The original 182x182 matrix size was cropped retaining the central 91x91 and 61x61 submatrices and the lower resolutions of 2.2x2.2x1.1 mm^3 and 3.3x3.3x1.1 mm^3 were respectively obtained. Voxels had a volume 4 or 9 times larger than the original voxels size, respectively. Note that the non-zero k-space values are identical for zero-filling and cropping, but the images reconstructed have different nominal resolutions.

We performed the preprocessing and the multivariate analysis pipeline as described for the original and zero-padded data, with the only difference that the number of selected voxels was matched in volume to the number of voxel selected in the original resolution (i.e., the number of voxels selected was divided by 4 or 9 for each feature selection level for the 2.2 mm and 3.3 mm resolution, respectively). Figure 2.8 shows that increasing the resolution led to a significant increase of decoding accuracies in case of vowel decoding, while no significant improvement of decoding accuracies was observed for speaker decoding. The p-values for the across-resolution comparison are reported in Table 2.1. These results are in agreement with those of the zero-padded datasets reported in the main text and therefore further support the conclusions of our study.

2.6.2 Control analyses for resolution and smoothing effects

In the ANOVA analysis testing the resolution and smoothing effect, we averaged the group accuracies across the levels of feature selection, therefore explicitly including the factors of resolution and smoothing, but not that of feature selection level. Although the accuracy values depend on the level of feature selection (i.e.,

the number of voxels), such dependency is similar for all accuracy curves as described in the discussion section. For this reason, averaging across the levels of feature selection does not introduce a bias. However, to directly address this issue, we performed two additional analyses: As a first control analysis, we performed a two-way repeated measures ANOVA separately for each level of feature selection. In Figure 2.9, we plot the resulting p-values for vowel and speaker decoding (left and right panel, respectively).

Note that the p-values were not corrected for multiple comparisons because the tests at different levels of feature selection are not independent, making such a correction not possible.

We observe that for both decoding dimensions, the main effect of smoothing is consistently below significance threshold for all levels of feature selection except for extremely low number of voxels. The main effect of resolution is consistently below significance threshold only for vowel decoding (excluding again extreme

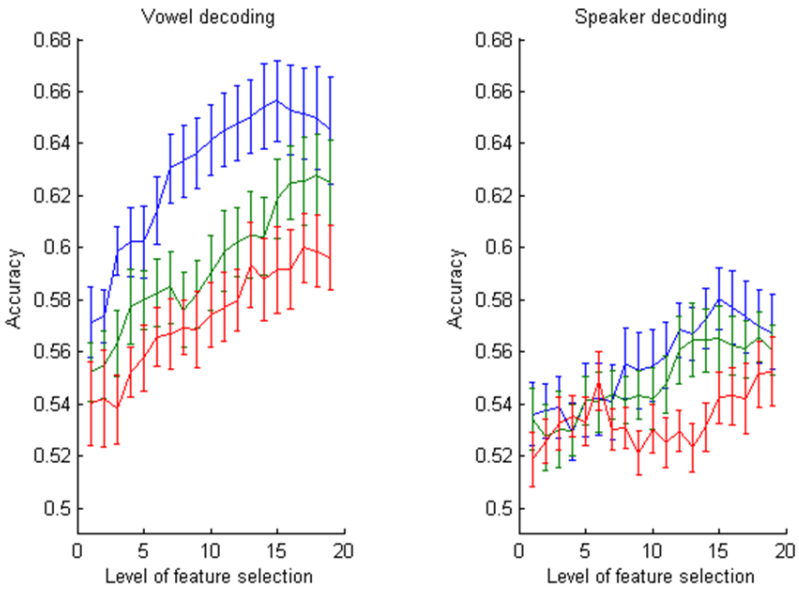

FIGURE 2.8: Resolution effect for lower nominal resolutions. Decoding accuracies are plotted as function of the feature selection levels for vowel and speaker decoding (left and right panel, respectively). The blue line represents the group level results for the original resolution, the green and red lines those for the lower resolutions (of 2.2x2.2x1.1 mm^3 and 3.3x3.3x1.1 mm^3, respectively). Error bars represent the standard errors.

TABLE 2.1: P-values of resolution effect for lower nominal resolution. The reported p-values refer to the comparison across resolutions illustrated in Figure 2.8. OR1 stands for original resolution at 1.1 mm isotropic, NR2 for nominal resolution at 2.2x2.2 mm^2 in-plane, NR3 for nominal resolution at 3.3x3.3 mm^2 in-plane. P-values were calculated via exact permutation test with $N = 10$ subjects; significant p-values are reported in bold.

	Vowel decoding	Speaker decoding
OR1 >NR2	**0.0059**	0.1670
OR1 >NR3	**0.0020**	0.1729
NR2 >NR3	**0.0039**	0.0928

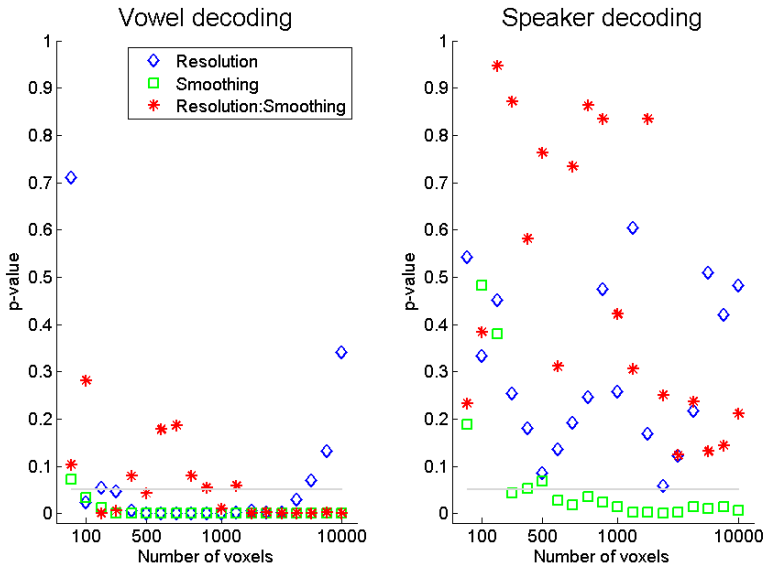

FIGURE 2.9: ANOVA separately for each level of feature selection. Main effect of resolution is in blue, main effect of smoothing in green and interaction between resolution and smoothing effect in red; the gray line represents the threshold of 0.05.

values of number of voxels selected). Finally, the interaction between resolution and smoothing effect is significant for the majority of the feature selection levels for only the vowel decoding. These results are therefore qualitatively confirming our findings.

As a second control analysis, we performed a three-way repeated measures ANOVA specifying the feature selection levels as 3^{rd} factor in addition to resolution and smoothing factors. We report the resulting F- and p-values in Table 2.2.

TABLE 2.2: Three-way repeated measures ANOVA. Significant p-values are reported in bold.

Effect	Vowel decoding		Speaker decoding	
	F	**p**	**F**	**p**
Resolution	10.14	**0.001**	1.58	0.233
Smoothing	13.65	**<0.001**	4.45	**0.002**
FeatureSelection	15.44	**<0.001**	23.78	**<0.001**
Resolution:Smoothing	2.67	**0.007**	1.02	0.433
Resolution:FeatureSelection	3.14	**<0.001**	1.57	**0.023**
Smoothing:FeatureSelection	7.00	**<0.001**	1.47	**0.004**
Resolution:Smoothing:FeatureSelection	2.22	**<0.001**	0.93	0.721

We observe that the main effect of resolution and smoothing as well as their interaction is confirmed. Moreover, the main effect of feature selection as well as its interaction with the other two factors is significant for both vowel and speaker decoding.

2.6.3 Validation of the BV likelihood method

In order to study the physiological origin of multivariate information in fMRI, we investigated the relationship between tissue type of voxels and their SVM weights. Tissue type classification was based on a ranking procedure that determines the likelihood of each voxel to belong to a gray matter or a large blood vessels region (therefore, here called *BV likelihood*). This procedure was based on the method introduced by Shmuel et al. (2010) and we further validated it by visually inspecting the obtained tissue type classification.

We built maps of the BV likelihood color-coding voxels from orange to yellow according to their rank (i.e., from low to high values indicating from "more gray matter" to "more large blood vessels", respectively).

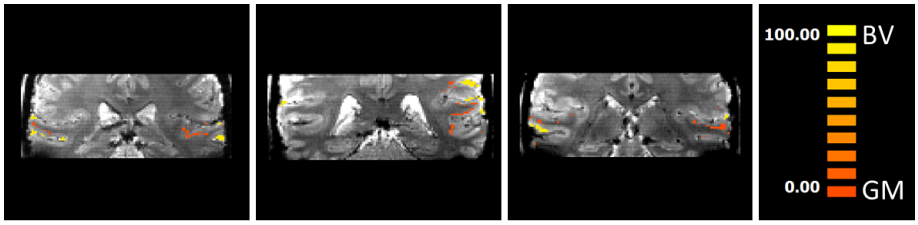

FIGURE 2.10: BV likelihood maps. Maps are thresholded to show only the 10% "more gray matter" voxels (in orange) and the 10% "more blood vessel" voxels (in yellow). Please, note that the maps included only the voxels that entered the classifier (because only for those voxels SVM weights were available), hence the very sparse appearance of the labelled voxels.

In Figure 2.10, we present the map of the BV likelihood values for three representative subjects thresholded to show only the 10% "more gray matter" voxels (in orange) and the 10% "more blood vessel" voxels (in yellow). BV likelihood maps are superimposed on the original EPI images, which offers clear structural contrast (with the gray matter being brighter than the white matter) thanks to the acquisition at ultra-high field. From this coronal view, we can observe that voxels labeled as "more blood vessels" are mainly localized on the surface of the cortex, while those labeled as "more gray matter" are localized more deeply in the cortex. Previous studies demonstrated that GE-BOLD changes in the surface layers are associated with extravascular BOLD effects due to superficial cortical vessels, while signal changes in deeper layers have microvasculature origin (Harel et al., 2006). The differential localization for the voxels classified at the two extremes of the BV likelihood is therefore in agreement with Harel et al. (2006) and confirms the validity of the used method.

Bibliography

N. Harel, J. Lin, S. Moeller, K. Ugurbil, and E. Yacoub. Combined imaging-histological study of cortical laminar specificity of fMRI signals. *Neuroimage*, 29(3):879–87, 2006.

A. Shmuel, D. Chaimow, G. Raddatz, K. Ugurbil, and E. Yacoub. Mechanisms underlying decoding at 7 T: ocular dominance columns, broad structures, and macroscopic blood vessels in V1 convey information on the stimulated eye. *Neuroimage*, 49(3):1957–64, 2010.

2.7 Supplementary material 2

2.7.1 Statistical testing by permutation at group level

Statistical testing at group level was performed by means of a permutation test. Although several comparisons were tested, the same algorithm was employed. In this section, therefore, we describe, using pseudocode, the general permutation procedure and subsequently illustrate how it was applied to each specific case tested in our study.

With this permutation test we compare two sets of decoding results, namely $V1$ and $V0$, each of them consisting of one accuracy curve (i.e., set of accuracy values at all nineteen feature selection levels) for each subject. The null hypothesis H0 is that the mean accuracy curves of $V1$ and $V0$ are the same, while the alternative is that the mean accuracy curve of $V1$ is larger than that of $V0$.

This non-parametric procedure was adopted in order to effectively account for the presence of correlation between accuracy values at different feature selection levels, as it is based on a distance measure that considers all the levels together. The rationale behind the procedure is to compare the observed distance between $V0$ and $V1$ with an empirical null distribution of such distance, obtained by exchanging, in all possible ways, the accuracy curves of each subject between the two sets $V1$ and $V0$.

The pseudocode below describes the algorithm used for the permutation test at group level.

```
% Variable definition

V1^(s) = accuracy curve 1 of subject s, considering all the feature
    selection steps f, therefore consists of 19 elements
V0^(s) = accuracy curve 0 of subject s, considering all the feature
    selection steps f, therefore consists of 19 elements
NumSubject = number of subjects (NumSubject = 10 in our study)
PermVector_i = i-th permutation (across all the possible 2^NumSubject
    permutations), consisting of 0s and 1s

% Group accuracy curves

⟨V1⟩ = mean of V1 across subjects, consisting of 19 elements in our study
⟨V0⟩ = mean of V0 across subjects, consisting of 19 elements in our study
⟨V1(f)⟩ = ⟨V1⟩ at feature selection level f
⟨V0(f)⟩ = ⟨V0⟩ at feature selection level f
```

% Metric to quantify the distance between the two group accuracy curves
% (where # counts the number of feature selection levels for which the
expression between brackets is true and the logistic sigmoidal
function remaps its argument between 0 and 1)

$$Distance(V1, V0) = \#\left(\langle V1\rangle > \langle V0\rangle\right) + logistic\left(\sum_{f=1}^{19}\left(\langle V1(f)\rangle - \langle V0(f)\rangle\right)\right)$$

% Distance between the ''real'' group accuracy curves

$$Dreal = Distance(V1, V0)$$

% Permutations

for perm = 1 : $2^{NumSubject}$

for s = 1 : $NumSubject$
if $PermVector_{perm}^{(s)}$ == 1
% swap accuracy curves for subject s
$PermV1^{(s)} = V0^{(s)}$
$PermV0^{(s)} = V1^{(s)}$
else
% do not swap accuracy curves for subject s
$PermV1^{(s)} = V1^{(s)}$
$PermV0^{(s)} = V0^{(s)}$
end
end
$\langle PermV1\rangle$ = mean of $PermV1$ across subjects, consisting of 19
elements in our study
$\langle PermV0\rangle$ = mean of $PermV0$ across subjects, consisting of 19
elements in our study

% Distance between the permuted group accuracy curves
$D(perm) = Distance\left(PermV1, PermV0\right)$

end

% p-value of the ''real'' group accuracy curves

$$p_{value} = \frac{\#(D \geq Dreal)}{2^{NumSubject}}$$

The function $Distance$ consists of two terms, where the first can be an integer

between 0 and 19 (centered on 9.5 under the null hypothesis H0), while the second is bound between 0 and 1 (centered on 0.5 under H0). Therefore, $Distance$ ranges between 0 and 20, and its expected value under H0 is 10. The first term reflects on how many feature selection levels $V1$ is higher than $V0$, whereas the second term weights the actual difference in mean between $V1$ and $V0$ across all the feature selection levels. Since the logistic function remaps these differences to the interval (0,1), the second term becomes relevant in the comparison between observed values and permuted values only when the first term is equal in both cases.

Importantly, the requirement of exchangeability under the null hypothesis, which is needed for the permutation test to be valid (Good, 2006), is satisfied since the subjects are independent and the metric defined by the function Distance is symmetrical under the null hypothesis (both terms in the summation are symmetric with respect to their expected values, under H0).

This algorithm was used for all group level permutation tests of our study, namely:

1. Permutation test to assess significance of a given curve versus chance, where:
   $V1^{(s)}$ = accuracy curve for vowel (speaker) decoding at resolution r (r = 1.1, 2.2 or 3.3 mm in our study) for subject s
   $V0^{(s)}$ = mean accuracy curve under the null hypothesis (obtained with single-subject permutation tests) for vowel (speaker) decoding at resolution r for subject s

2. Effects of resolution, where:
   $V1^{(s)}$ = accuracy curve for vowel (speaker) decoding at resolution $r1$ for subject s
   $V0^{(s)}$ = accuracy curve for vowel (speaker) decoding at resolution $r2$ (with $r2 > r1$) for subject s

3. Difference of effect of resolution across decoding tasks, where:
   $V1^{(s)}$ = difference between accuracy curves at resolutions $r1$ and $r2$ (with $r2 > r1$) for vowel decoding for subject s
   $V0^{(s)}$ = difference between accuracy curves at resolutions $r1$ and $r2$ (with $r2$, $r1$ same resolutions as in $V1^{(s)}$ definition) for speaker decoding for subject s

Bibliography

P. I. Good. *Permutation, parametric, and bootstrap tests of hypotheses*. Springer Science & Business Media, 2006.

Chapter 3

Tonotopic maps in human auditory cortex using ASL

Based on:

Gardumi, A., Ivanov, D., Havlicek, M., Formisano, E., and Uludağ, K. (2016). *Tonotopic maps in human auditory cortex using Arterial Spin Labeling*. Human Brain Mapping (In press)

3.1 Abstract

A tonotopic organization of the human auditory cortex (AC) has been reliably found by neuroimaging studies. However, a full characterization and parcellation of the AC is still lacking. In this study, we employed pseudo-continuous arterial spin labeling (pCASL) to map tonotopy and voice selective regions using, for the first time, cerebral blood flow (CBF). We demonstrated the feasibility of CBF-based tonotopy and found a good agreement with BOLD signal-based tonotopy, despite the lower contrast-to-noise ratio of CBF. Quantitative perfusion mapping of baseline CBF showed a region of high perfusion centered on Heschl's gyrus and corresponding to the main high-low-high frequency gradients, co-located to the presumed primary auditory core and suggesting baseline CBF as a novel marker for AC parcellation. Furthermore, susceptibility weighted imaging was employed to investigate the tissue specificity of CBF and BOLD signal and the possible venous bias of BOLD-based tonotopy. For BOLD only active voxels, we found a higher percentage of vein contamination than for CBF only active voxels. Taken together, we demonstrated that both baseline and stimulus-induced CBF are an alternative fMRI approach to the standard BOLD signal to study auditory processing and delineate the functional organization of the auditory cortex.

3.2 Introduction

Neuroimaging techniques allow for the non-invasive investigation of the functional organization of the auditory cortex (AC) in the human brain: In agreement with previous animal studies (e.g., cats and primates; Merzenich and Brugge, 1973; Merzenich et al., 1973), functional magnetic resonance imaging (fMRI) studies have found a tonotopic organization of the human early AC (Formisano et al., 2003; Talavage et al., 2004; Humphries et al., 2010; Woods et al., 2010; Da Costa et al., 2011; Langers and van Dijk, 2012). Human early AC tonotopy has been consistently described as two frequency gradients composing a high-low-high preferred frequency pattern located in and around the Heschl's gyrus (HG). However, its extent, orientation, and finer details are still a matter of debate. In the visual cortex, the early visual areas are parcellated by means of stimuli varying within the two dimensional extent of the visual field, called retinotopy. In contrast, in the auditory domain, a second stimulus dimension is so far missing that would enable delineation of the borders of early auditory areas.

Within the high-low-high frequency map, the more posterior/medial frequency gradient is considered the human homolog of monkey area A1, while the more anterior/lateral frequency gradient is considered the human homolog of monkey area R. These two areas (gradients) together are considered to define the human *primary auditory cortex* (PAC, also called *auditory core*) (Baumann et al., 2013; Moerel et al., 2014; Saenz and Langers, 2014). However, the exact extent of the PAC remains ambiguous: Some studies consider the main high-low-high frequency gradients to be entirely included in the PAC (Da Costa et al., 2011), while other studies attribute part of them to auditory belt regions (Talavage et al., 2004; Humphries et al., 2010). Moreover, three main theories about the orientation of the PAC coexist: The *classical* interpretation locates the PAC along HG and finds its foundations in cytoarchitectonic studies. The *orthogonal* interpretation locates the PAC across HG assuming in between-species homology with monkeys (for which PAC was shown to run parallel to the superior temporal gyrus, STG). The *oblique* interpretation is a more recent theory proposing an oblique orientation of PAC with respect to HG (Baumann et al., 2013). Finally, additional tonotopic gradients outside of the HG are often observed, especially at high spatial resolution and single-subject level (Da Costa et al., 2011; Herdener et al., 2013; Moerel et al., 2013).

Besides tonotopy, other functional organizations of AC related to sound properties have been investigated, such as tuning width and periodicity preference (Barton et al., 2012; Moerel et al., 2012; Herdener et al., 2013; Moerel et al., 2013). Moreover, also higher order functional areas have been studied: Specifically, regions showing differential response for vocalizations with respect to other sound categories (tones, objects, environmental noises, etc.) have been reliably observed and are typically referred to as *voice sensitive areas* (Belin et al., 2000).

In addition to tonotopic gradients, anatomical landmarks are utilized to identify PAC. Recently, "myelin" imaging using MRI has been proposed for that purpose (De Martino et al., 2015). Primary sensory areas, such as early visual, somatosensory and auditory cortex, have been hypothesized to have higher myelin content than the surrounding brain areas, and they are detectable with T1 and T2* MRI contrasts (Sigalovsky et al., 2006; Bock et al., 2009; Geyer et al., 2011; Glasser and Van Essen, 2011; Cohen-Adad et al., 2012; Dick et al., 2012; Sereno et al., 2013; De Martino et al., 2015). Alternatively, susceptibility weighted imaging (SWI) can be used to probe iron and myelin differences in the cortex and, additionally, to locate veins. SWI is a recent MR technique detecting susceptibility differences in the brain (Reichenbach et al., 1997; Haacke et al., 2004). SWI combines magnitude and phase information of the complex T2* weighted images to enhance the contrast of paramagnetic substances, such as deoxygenated hemoglobin and iron, with respect to the surrounding diamagnetic tissue.

FMRI studies investigating the functional organization of AC have so far employed the blood oxygenation level-dependent (BOLD) effect as an indirect measure of neural activity. Although the BOLD signal is the standard contrast for fMRI, it has some limitations in terms of spatial specificity and of being a quantitative marker of neuronal activity. The BOLD signal arises from the combined changes of oxygen metabolism ($CMRO_2$), cerebral blood flow (CBF), and cerebral blood volume (CBV) in response to neural activity modulations (Logothetis, 2008). It has been shown that the overall changes in the oxygenation spread from the location of the neural activity into the draining veins. At low magnetic field strength (e.g., 1.5 and 3 T) and for both gradient-echo and spin-echo sequences, the measured BOLD signal mostly originates from draining veins (Uludag et al., 2009). Consequently, the spatial specificity of the BOLD signal is biased by the presence of draining veins causing signal blurring and possible displacement from the actual site of neural activity (Ugurbil et al., 2003).

Besides the BOLD signal, alternative fMRI acquisition techniques exist to study

brain function. For example, arterial spin labeling (ASL) techniques measure absolute cerebral blood flow (CBF) in addition to the BOLD signal. ASL allows to quantify both stimulation-induced and baseline CBF as an absolute marker of the physiological state of the tissue and its changes. Studies have shown that CBF, compared to the BOLD signal, is more spatially localized to neural activity, has lower intersubject variability and is more reproducible over time (Aguirre et al., 2002; Wang et al., 2003; Tjandra et al., 2005). However, CBF signal has a lower SNR and its quantitative estimation is challenged by MRI acquisition confounds such as transit delay, magnetization transfer and relaxation time effects. Nevertheless, the post-labeling delay (PLD) interval (i.e., the time interval occurring between blood labeling and image acquisition, which is necessary for the labeled blood to flow from the tagging location to the imaging slab) represents an inherently silent gap in the ASL MR sequence, allowing therefore auditory stimulation in the absence of scanner noise and making ASL particularly attractive for auditory fMRI studies.

In this paper, we aim to assess — for the first time — the feasibility of tonotopic mapping of the human auditory cortex with ASL fMRI, and to compare BOLD and CBF based tonotopies. Additionally, we evaluate whether baseline quantitative CBF and SWI provide information on the location of the PAC in addition to anatomical landmarks and myelination.

3.3 Material and methods

3.3.1 Subjects

Twelve healthy volunteers (six females, age range 25–33) with normal hearing took part in this experiment. Written informed consent was obtained from all participants according to the approval of the study protocol by the Ethical Committee of the Faculty of Psychology and Neuroscience, Maastricht University.

3.3.2 Stimulus design and presentation

The stimulus design employed in this work (if not stated otherwise) was adapted from the study by De Martino et al. (2013, cfr. Experiment 3), which investigated tonotopy in the inferior colliculus and auditory cortex using BOLD fMRI. The stimuli consisted of amplitude modulated (8 Hz, modulation depth 0.95, length

0.8 s, sample rate 44.1 kHz) tones created in MATLAB with 8 carrier frequencies logarithmically spaced between 0.180 and 7.091 kHz (see Figure 3.1B). Each of these 8 frequencies represents one stimulus condition and, for the rest of the paper, we will refer to them as *center frequencies*. In order to introduce variability within each stimulus condition, for each center frequency two additional sounds were created with a frequency difference of $\pm 10\%$ in logarithmic scale. A total of 24 tones were therefore created (center frequencies are highlighted in bold): 168,**180**,193, 284,**304**,326, 480,**514**,551, 811,**869**,931, 1370,**1469**,1574, 2316,**2482**,2661, 3915,**4196**,4497, 6616,**7091**,7600 Hz. Sound onset and offset were ramped with a 10 ms linear slope and the sound energy (calculated as root mean square) was equalized.

In one set of experiments (subjects nr 1 to 6), stimulus conditions were presented in blocks of six TRs with one sound per TR (= 3 s) (same as in De Martino et al., 2013). Each stimulus block was followed by a resting block of six TRs with no auditory stimulation. In another set of experiments (subjects nr 7 to 12), the stimulus duration was reduced to four TRs and the rest condition increased to eight TRs. Stimulus conditions were presented in randomized order and repeated twice per run. Six runs were acquired for each subject, for a total of 96 stimulus blocks.

An additional voice localizer run was acquired (except for subject 1) using the same block design as above. For this run, however, the auditory stimuli used (adapted from Bonte et al., 2013) were 1.0 s long and consisted of vocal sounds (both speech and non-speech), other natural sounds (musical instruments, environmental and tool noises, and animal cries), or amplitude-modulated tones (8 Hz, frequency between 0.3 and 3.0 kHz). The run included 8 blocks of each of these 3 categories for a total of 304 TRs and $\sim$15 min duration.

During all functional runs, subjects were asked to fixate a cross in the center of the screen and passively listen to the sounds.

3.3.3 MRI acquisition

Measurements were performed on a 3T Prisma Siemens scanner using a 64-channel head coil. The functional runs were acquired using pseudo-continuous ASL (pCASL) with 2D single-shot echo-planar imaging (EPI) readout (TR 3 s, TE 13 ms, voxel size 2.5 mm isotropic, 19 slices, labeling duration 1.2 s, post labeling delay 1.2 s, partial Fourier 7/8, GRAPPA 2; Dai et al., 2008). Labeling duration

and PLD were deliberately chosen shorter than recommended by the ASL white paper (Alsop et al., 2015), whose recommendations are tailored in view of whole brain baseline perfusion measures. These choices allow us shorter TR and thus improved sampling of the hemodynamic response. At the same time, PLD was chosen long enough to allow CBF quantification in the auditory cortex, which is adequately perfused by the labeled blood after a PLD of 1.0 s (Donahue et al., 2014; Mezue et al., 2014).

This sequence is acoustically characterized by a tagging module and EPI-train presenting a power spectrum with contribution mostly from frequencies below 3-4 kHz (peak frequencies were 990.5, 1981.0, and 2993.0 Hz for the tagging module and 925.9, 1163.0, 1809.0, and 2713.0 Hz for the EPI-train; see Figure 3.1) and relative loudness (estimated as the ratio between the RMS values of the tagging module and the EPI readout waveforms) of 0.57. Functional images were acquired while participants passively listened to the tones presented in a block design. During the stimulus blocks, the sounds were presented in the 1.2 s PLD interval (see Figure 3.1A). At the beginning of the fMRI session, the stimuli were subjectively equalized for loudness, and the overall volume of the stimuli was adjusted to a comfortable intensity level. The stimuli were presented via MR-compatible earphones (Sensimetrics S14, Malden, MA, USA). After the experiment, all subjects reported a clear hearing of the stimuli.

In order to allow absolute quantification of CBF (see below), an M_0 image was acquired using the same pCASL sequence as described above but with the TR value increased to 20.0 s.

A susceptibility weighted image was acquired with an in-plane resolution of 0.5x0.5 mm^2 and a slice thickness of 1.0 mm (TR 28 ms, TE 20 ms, Flip angle 15deg, GRAPPA 2, matrix size 384x312x52; Reichenbach et al., 1997; Haacke et al., 2004). Finally, a high-resolution (1.0 mm isotropic) anatomical image was acquired using an MPRAGE sequence (TR 2.4 s, TE 2.18 ms, TI 1040 ms, Flip angle 8deg, GRAPPA 2, matrix size 224x224x192; Mugler and Brookeman, 1990).

3.3.4 Preprocessing

Functional data were preprocessed in BrainVoyager QX (Version 2.8.4.2645, 64-bit, Brain Innovation, Maastricht, The Netherlands): The functional runs were motion corrected and realigned to the first volume of the fourth functional run (i.e., the image approximately at the midpoint of the session and closest in terms

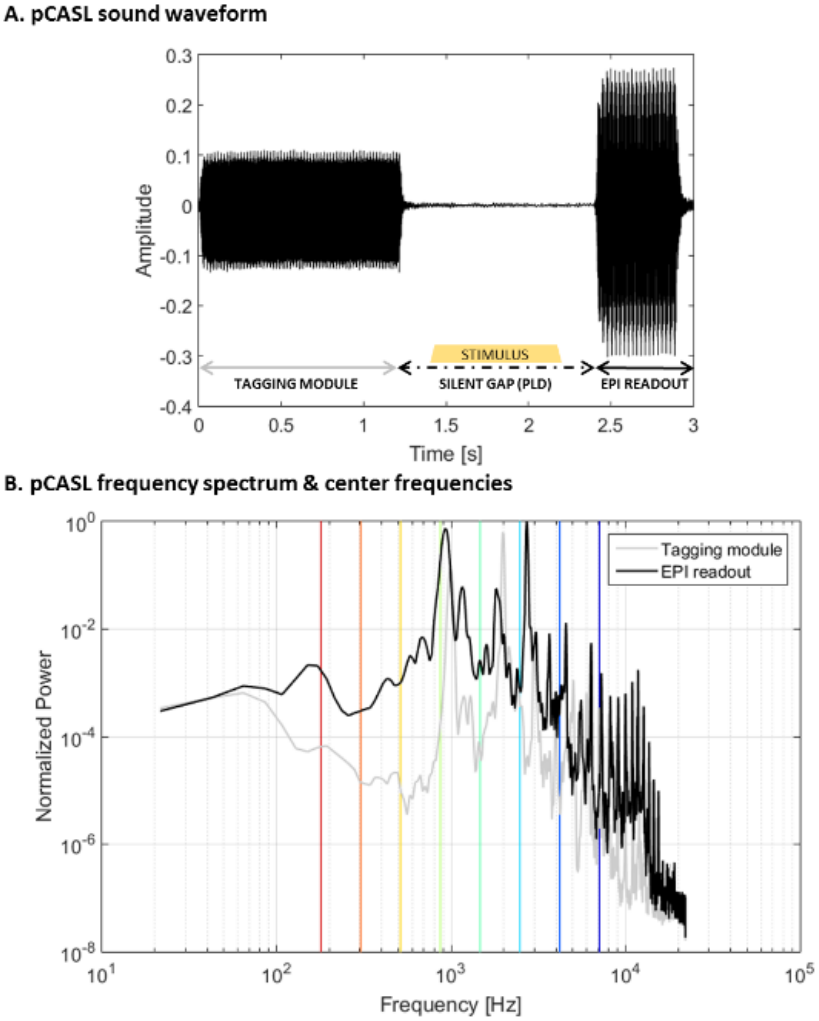

FIGURE 3.1: Sound waveform (A) and frequency spectrum (B) of pCASL with 2D single-shot EPI acquisition. In Panel A, the different components of a pCASL TR are illustrated: the tagging module, the silent gap (PLD; in which a stimulus sound may occur following the experimental design), and then the EPI-train readout. In Panel B, the gray and black lines show the frequency spectrum of the tagging module and the EPI-train readout, respectively. The vertical bars indicate the central frequencies of the stimuli and are color-coded in the same manner as the tonotopic maps (i.e., from red-to-green-to-blue for low-to-medium-to-high frequency).

of acquisition order to the SWI image, which was acquired between the third and fourth functional run). 3D motion correction was performed using the 'Trilinear/sinc interpolation' option, i.e., trilinear interpolation is used during the motion detection step and sinc interpolation for the actual motion correction (spatial transformation) step. Functional ASL, M_0 and SWI image were coregistered to the anatomical image using a gradient-based alignment with six-parameter (i.e., three translation and three rotation parameters) rigid body transformation.

The anatomical image was transformed into Talairach space in order to employ the automatic segmentation implemented in BrainVoyager. The results of the automatic segmentation were visually inspected and manually corrected, whenever necessary. The obtained white-gray matter boundary was reconstructed to produce a 3D folded cortical surface for each subject. This 3D cortical representation was used for group alignment purposes to the aim of both computation and visualization (after inflation) of group maps. The group cortical surfaces were produced taking into account the individual subject's cortical curvature via the 'moving target group averaging approach' of the cortex-based alignment (CBA) procedure (Fischl et al., 1999; Goebel et al., 2006). Anatomical masks of the temporal lobe and the primary auditory cortex were manually drawn in the common CBA space based on (Baumann et al., 2013; Bonte et al., 2013; Kim et al., 2000, see Figure 3.2A).

3.3.5 BOLD and CBF time courses

The BOLD and CBF time courses were calculated from the motion corrected ASL time course using the BrainVoyager ASL Perfusion Volume Data Processing plugin performing surround averaging and subtraction: The ASL time course is separated in a control and label time-series. Each time-series is temporally interpolated to obtain BOLD and CBF data points at each original TR. The subtraction of the interpolated label from the interpolated control time-series yields the CBF time course, while the addition of the two interpolated time-series yields the BOLD time course (Liu and Wong, 2005).

Both BOLD and CBF time-series were transformed from the functional space to the anatomical space of each individual subject by applying the coregistration transformation previously calculated and the data was slightly upsampled to an isotropic resolution of 2 mm (i.e., half the anatomical resolution). Finally, both time-series were spatially smoothed in the volume space with a 2 mm FWHM 3D

Gaussian kernel (Gardumi et al., 2016) and temporally high-pass filtered removing linear and low frequency non-linear drifts up to 3 cycles per time course.

In order to evaluate the signal quality of the CBF and BOLD time-series, two measures of signal-to-noise ratio (SNR) were employed: the temporal signal-to-noise ratio (tSNR) and the contrast-to-noise ratio (CNR). The tSNR was calculated as the mean of the time course divided by its standard deviation and averaged across all the voxels in the temporal mask of AC (see Figure 3.2A). The CNR, which represents a quantity more closely linked to the functional sensitivity of the data, was calculated as the ratio between the standard deviation of the activation response and the standard deviation of the noise (Welvaert and Rosseel, 2013). In the CNR calculation, only active (as defined via a GLM analysis, see following section) voxels were included. The CNR was computed for all center frequencies together and for each of them separately. A two-way repeated measure ANOVA, performed with contrast (CBF or BOLD signal) and center frequency (i.e., the eight stimulus conditions) as factors and a significance level α of 0.05, was used to assess differences in CNR.

In order to investigate the consistency of CBF and BOLD signal changes across subjects, we computed the coefficient-of-variation (COV), defined as the standard deviation divided by the mean of the percent signal change across the subjects and expressed in percentage. For each subject, the percent signal change was calculated as the ratio between the amplitude of the mean evoked response and the temporal mean of the time course using the same number of active voxels for CBF and BOLD signal.

3.3.6 Computation of the BOLD- and CBF-based tonotopic maps

BOLD- and CBF-based tonotopic maps were independently calculated following the same two-step procedure:

First, a general linear model (GLM) analysis was performed using one predictor for each of the eight stimulus conditions (i.e., each of the eight center frequencies) to model the respective BOLD/CBF response. The predictors were built convolving a box car function representing the stimulus block with a canonical (double gamma) hemodynamic response function (HRF) in order to account for the hemodynamic response delay. For each run, one constant predictor and the six parameters estimated by the motion correction algorithm were included in the GLM as confound predictors. The obtained BOLD/CBF statistical maps were

thresholded (uncorrected t-value > 2 for the contrast: all center frequencies $>$ baseline) to select only active voxels entering the 2^{nd} step.

Second, for each active voxel, its BOLD signal/CBF preferred frequency was defined as the one having the highest β-value among the BOLD signal/CBF predictors. The final single-subject BOLD signal/CBF tonotopy was obtained by color-coding each voxel according to its preferred frequency: with red-to-green-to-blue coding for low-to-medium-to-high frequencies taking into account the logarithmic spacing of frequency in the stimuli. For visualization, the maps were projected on the inflated cortical surfaces. The color-coded single-subject maps were transformed into the common CBA space and averaged to obtain a group tonotopic map (Formisano et al., 2003; Langers et al., 2007; Humphries et al., 2010; Da Costa et al., 2011; Moerel et al., 2012; De Martino et al., 2013; Herdener et al., 2013).

Results from the GLM analysis performed as first step of the tonotopy computation were used also to assess the overall response to the sound stimuli as measured by CBF and BOLD signal. To evaluate the extension of activation in the auditory cortex and its spatial consistency across subjects, probabilistic maps were computed by calculating at each vertex of the common CBA space the relative number of subjects having significant activity at that spatial location. For this analysis, a threshold of $Q(FDR) < 0.05$ (contrast: all center frequencies $>$ baseline; FDR standing for false discovery rate) was used. The difference in number of voxels significantly active for CBF and BOLD signal was tested using a two-tailed t-test (with significance level α of 0.05).

3.3.7 Comparison of BOLD- vs CBF-based tonotopic maps

BOLD and CBF tonotopies were compared by calculating their spatial correlation at the single-subject level in the native anatomical space of each subject and at group level in the CBA surface space. The significance of the correlation at single-subject level was evaluated by estimating the null-distribution via permutation test (N_perm = 1000). For each iteration of the permutation test, the spatial correlation was computed between the original BOLD tonotopy and the permuted CBF tonotopy obtained by randomly permuting the preferred frequencies across the voxels in the tonotopic map. Here and in the rest of the paper, unless specified, by the term correlation, we refer to Pearson's correlation calculated between the two maps represented in vectors.

3.3.8 Baseline CBF quantification

A quantitative perfusion (CBF) map was estimated from the ASL data using the model proposed by the ASL white paper (Alsop et al., 2015). The assumption of this model are: complete bolus delivery to the target tissue (i.e., PLD $>$ ATT for pCASL, where ATT stands for arterial transit time); no venous outflow of labeled blood water (which is generally valid in humans); and T1 relaxation time of labeled spins to be the same as T1 of blood (at 3T, these values are similar and therefore errors are negligible (Cavusoglu et al., 2009)). Given that the assumptions are satisfied in our study, quantitative CBF was calculated in each voxel using the following equation:

$$qCBF = \frac{6000 \cdot \lambda \cdot \beta \cdot e^{\frac{PLD}{T1_{blood}}}}{2 \cdot \alpha \cdot T1_{blood} \cdot M_0 \cdot \left(1 - e^{-\frac{\tau}{T1_{blood}}}\right)} \, [ml/100g/min] \tag{3.1}$$

where β is obtained by estimating the full GLM ASL model (for details see Mumford et al., 2006; Hernandez-Garcia et al., 2010). In this model, β is a scaling parameter of the baseline CBF predictor (constructed as an alternation of -0.5 and 0.5) and is proportional to the label and control signal difference. Since CBF and BOLD signal changes related to activation are also modeled in this GLM approach, the β estimate of baseline CBF is not biased by signal changes induced by the task auditory stimulation. Other parameters in Equation 3.1 are: M_0 representing the signal intensity of the M_0 image; PLD the post labeling delay (adjusted for each slice according to the slice acquisition order); τ the label duration; $T1_{blood}$ the longitudinal relaxation time of blood in seconds ($T1_{blood}$ = 1.650 s at 3T; Lu et al. (2004), Zhang et al. (2013)); λ the brain/blood partition coefficient in ml/g (λ = 0.9 ml/g); and α the labeling efficiency (α = 0.85 for pCASL). Finally, the factor 6000 converts the units from $1/$s to ml/100g/min, which is the commonly used physiological unit for quantitative CBF.

The resulting quantitative CBF map was used to calculate single-subject gray matter (GM) perfusion values by averaging the CBF values across all voxels included in a GM mask defined on the base of the individual anatomical segmentation and intersected with the ASL imaging slab. The group perfusion map was obtained as the average of the single-subject perfusion maps after registration in the CBA space.

3.3.9 Voice selective regions

The BOLD and CBF time-series calculated from the ASL voice localizer run were projected on the individual surface space and, independently, analyzed with a GLM including one predictor for each category (vocal sounds, other natural sounds and tones). The predictor was built as the convolution of the box car representing the stimulus block with a canonical (double gamma) HRF. β estimates of all predictors were calculated at each vertex of the cortical surfaces. Then, the individual β estimates were projected in the CBA surface space and a second level (i.e., multi-subject) GLM was performed taking into account the variability across the subjects (random effects group analysis, RFX). The obtained CBF and BOLD statistical maps were thresholded (uncorrected t-value > 2; contrast: all voice localizer sounds $>$ baseline) to select only vertices showing stimulus-induced activity. Such vertices were included in the computation of the contrast vocal sounds $>$ (other natural sounds + tones)/2. In this manner CBF and BOLD unthresholded voice selective maps were obtained. Their (dis)agreement was assessed computing the correlation between the two maps. The approach presented here intentionally avoids to statistically threshold the maps at significance level in order to compare CBF and BOLD-based voice selective maps circumventing the issue of lower SNR for CBF compared to BOLD signal. Their comparison through correlation is a valid approach, but unthresholded maps have to be interpreted with caution. In the Supporting Information, the comparison of CBF and BOLD voice selective regions was performed also using statistically thresholded maps, with qualitatively similar results (see Supporting Information 3.7.6 and 3.7.7).

In order to quantitatively compare our results with previous findings using standard BOLD fMRI, we computed the correlation between the BOLD unthresholded voice selective map obtained in this study and that made available at `http://neurovault.org/collections/33/` by Pernet et al. (2015). The correlation was calculated after transforming the latter from MNI space to our group-specific CBA space and considering only the common vertices between the two maps.

3.3.10 Vein masks from susceptibility weighted images

Vein masks were generated from SWI in order to assess the tissue specificity of the BOLD and CBF activation signal. SWI takes advantage of the complementary information of T2* weighted magnitude and phase images (Reichenbach et al., 1997;

Haacke et al., 2004; Rauscher et al., 2006). Because of their paramagnetic properties, venous vessels look dark on magnitude images and take negative values in the phase image. SWI uses the phase information to further suppress the magnitude intensity of venous vessels and therefore enhance their detection in the final SW image. A sliding minimum intensity projection (mIP) over two subsequent slices of the SW image was performed in order to profit from vessel connectivity while preserving the local information across the slice direction to the level of the resolution of the functional maps (i.e., 2 mm). Vein masks were created by binarising the mIP-SW image with the value 1 assigned to voxels having a SWI value lower than 1/5 of the maximum SWI value and the value 0 otherwise. The vein mask was coregistered to the individual anatomy applying the coregistration parameters estimated between magnitude image of the SWI and MPRAGE image of the anatomy. Finally, the coregistered vein mask was downsampled to an isotropic resolution of 2.0 mm to match the resolution of the functional data (Harmer et al., 2012). The resulting vein mask (see Supporting Information Figure 3.5) was visually inspected by overlaying it on the original SW image and the T1 weighted anatomy. Note that the obtained vein mask may include CSF voxels given that also CSF signal is suppressed in SWI images; however, this does not represent a concern for our analyses since we used SWI information only for voxels detected as active by GLM of CBF and/or BOLD signal and therefore most likely consisting of tissue, vessels and/or CSF containing vessels.

The fraction of active voxels, as previously defined by a GLM analysis of the BOLD and CBF signal, labelled as vein voxels in the vein mask was determined. Then, we investigated whether the presence of a vein biasing more strongly the BOLD signal than the CBF signal could explain the partial mismatch between the BOLD and CBF tonotopies. To that end, we calculated the correlation between BOLD and CBF tonotopies splitting the voxels according to the vein mask labelling.

3.4 Results

3.4.1 Activation in auditory cortex

Figure 3.2A shows the two anatomically defined masks used in this study: in light blue, the temporal cortex mask including the superior temporal plane (STP), STG, superior temporal sulcus (STS) and middle temporal gyrus (MTG); in pink the

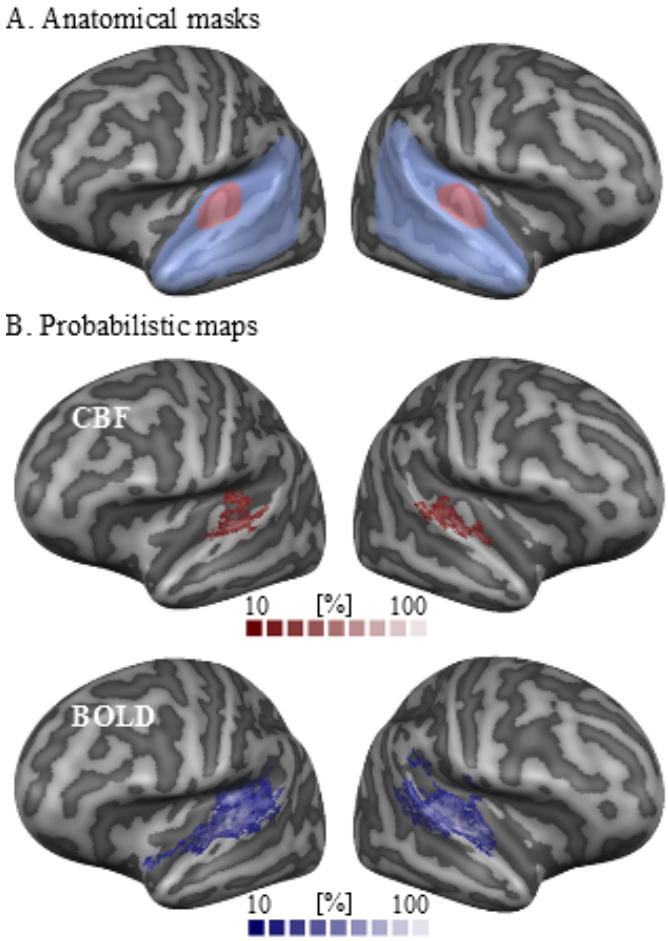

FIGURE 3.2: Anatomical masks and probabilistic maps of activation. Panel A shows the anatomically defined temporal (in light blue) and primary auditory cortex (in pink, overlapped on it) masks. Panel B shows the probabilistic map (from 10 to 100% overlap across subjects) of activation in response to tones. Activation maps from the CBF time-series are shown in red (top row), while those from the BOLD time-series in blue (bottom row).

primary auditory cortex mask including HG and the areas immediately anterior and posterior to it. Masks were drawn in the common CBA space based on (Kim et al., 2000; Baumann et al., 2013; Bonte et al., 2013).

Tones elicited a robust activation in the auditory cortex as measured by both BOLD and CBF signals and shown in Figure 3.2B by a probabilistic map of the

contrast between all stimulus conditions and baseline. The activated areas included HG, planum temporale (PT), planum polare (PP), STG, and STS. BOLD activation clusters were more widespread than CBF ones, as expected because of the lower SNR of CBF signal compared to the BOLD signal. Thresholding each single-subject statistical map with $Q(FDR) < 0.05$, the resulting number of active voxels was significantly higher for the BOLD signal with respect to CBF signal (2036±274 [range 414-3303] for BOLD signal and 729±108 [range 140-1458] for CBF, reported as mean ± standard error across the subjects and [range min-max value]; $p < 0.001$, two-tailed t-test; see Supporting Information Figure 3.6 for a boxplot of the value distributions). The tSNR, calculated averaging across all the voxels in the temporal cortex mask of AC, was 57.6 ± 1.4 for BOLD signal and 2.3 ± 0.1 for CBF (where the tSNR values are reported as mean ± standard error calculated across the subjects). More closely linked to the functional sensitivity of the data, the CNR of the two time-series resulted in a value of 0.206 ± 0.014 for the BOLD signal and 0.130 ± 0.004 for CBF (reported as mean ± standard error across the subjects). CNR was also calculated separately for each center frequency and a two-way repeated measure ANOVA was performed with contrast (CBF or BOLD signal) and center frequency (i.e., the eight stimulus conditions) as factors. Results showed a significant main effect for the contrast ($p < 0.001$), but no significant effect for the center frequency nor significant interaction between the two factors.

The average percent signal change was $1.53 \pm 0.12\%$ for the BOLD signal and $16.50 \pm 1.19\%$ for CBF signal (reported as mean ± standard error calculated across the subjects), resulting in a COV of 27.73% and 24.99%, respectively.

3.4.2 CBF and BOLD tonotopies

Tonotopic maps for CBF and BOLD signal are shown in Figure 3.3A (top and bottom row, respectively). The CBF tonotopy presented two reversed spatial gradients of preferred frequencies located on HG: preferred low frequencies (in red) were located on the central part of HG and preferred high frequency (in blue) medially and posteriorly to it, forming a gradient pattern of high-low-high frequency. Additional gradients were located in the surrounding areas. More specifically, clusters of low frequencies were identifiable on the middle part of the STG lateral and anterior to HG, and on the posterior STG. The tonotopic patterns are similar across the left and right hemispheres. The overall layout and the spatial

arrangement of the tonotopic gradients described for the CBF tonotopy is in good (qualitative) agreement with those in the BOLD tonotopy (in the corresponding panel of Figure 3.3A, the two reversed gradients forming the high-low-high frequency pattern are indicated by white double arrows and the additional low frequency clusters by black single arrows), and similar to maps shown in previously published BOLD signal studies (see e.g., Formisano et al., 2003; Humphries et al., 2010; Da Costa et al., 2011; Moerel et al., 2012). However, in the CBF tonotopy, extreme low or high preferred frequency values are less represented than in the BOLD tonotopy. A few limited mismatches between CBF and BOLD signal tonotopy are highlighted in Figure 3.3A by white single arrows.

The correlation between the CBF and BOLD tonotopic maps in the individual volume space was significantly above chance for all subjects but one (i.e., 11 out of 12 subjects; $p < 0.01$ as assessed by permutation test) resulting in a mean correlation of 0.15 ± 0.06.

The spatial correlation between the BOLD and CBF group tonotopic maps was calculated in the group-aligned surface space and resulted in a value of 0.45. Such correlation increased to a value of 0.67 when restricting the computation to the vertices within an anatomically defined mask of primary auditory cortex (PAC; see Figure 3.2A, pink mask). In this ROI, the right hemisphere showed a slightly higher correlation value than the left one (0.72 and 0.65, respectively).

3.4.3 Perfusion map and parcellation of the primary auditory cortex

A gray matter baseline perfusion value of 54 ± 2 ml/100g/min was obtained as mean and standard error across the subjects. Supporting Information Figure 3.7 shows a multi-slice view of the quantitative perfusion map for each subject. Figure 3.3B shows the group perfusion map, the histograms of the gray matter perfusion values at vertex level in the left and right hemisphere, and a zoomed view of the temporal lobes after thresholding ($value > 68$ ml/100g/min) in order to highlight the region(s) with higher baseline perfusion. (Note that due to the limited brain coverage of the ASL acquisition, there are no CBF values detected for the top part of the cortex.) In left and right AC, a relatively homogeneous region with high perfusion was centered on HG and extended posteriorly and medially to it. These areas corresponded in both hemispheres to the main high-low-high frequency pattern observed in the previously computed tonotopic maps and to

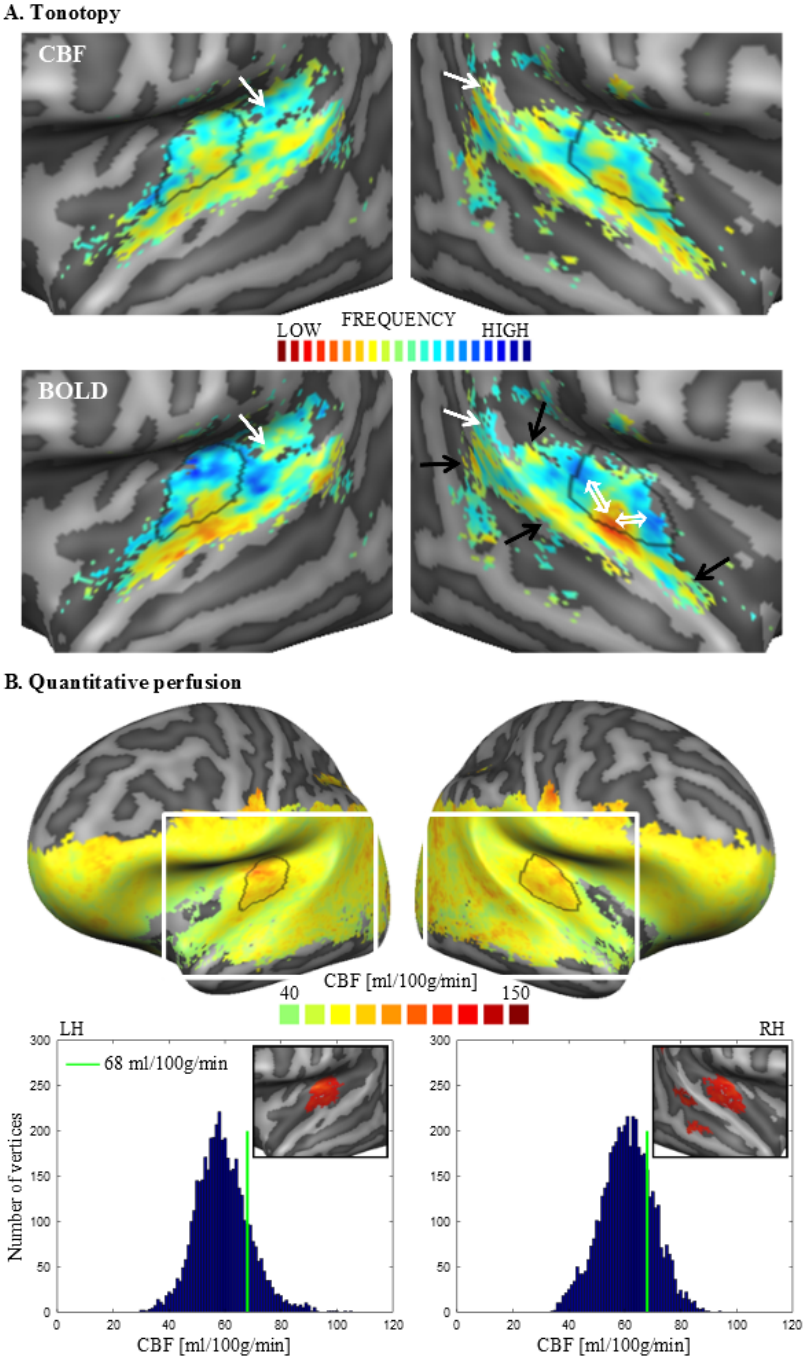

FIGURE 3.3: (A) Group tonotopy and (B) quantitative baseline perfusion maps.

Panel A shows the group tonotopy obtained from CBF (top row) and the BOLD signal (bottom row). Overlaid to the right BOLD map, two white double arrows show the main gradients composing the V-shaped frequency pattern of PAC and the black single arrows show clusters of low frequencies, which might belong to additional gradients outside PAC. Given the symmetry between hemispheres and the good agreement between CBF and BOLD tonotopy, such guiding symbols are presented only in one panel. In both CBF and BOLD tonotopic maps, regions presenting a mismatch between the two maps are indicated by white single arrows. Panel B shows the group quantitative perfusion map and the histograms of the perfusion values separately for the two hemispheres (LH and RH stand for left and right hemisphere, respectively). In the plots, the green bar indicates the threshold of 68 ml/100g/min (corresponding to the 80^{th} percentile of the baseline perfusion in the temporal mask). The white rectangle overlaid to the hemispheres shows the region of the close-up view used for Panel A and the inserts in the histogram plots below. The inserts show a close-up view of the thresholded perfusion map ($CBF > 68$ ml/100g/min; masked with the anatomically defined temporal mask (Figure 3.2A, blue mask) and 25 mm^2 cluster size threshold), in which one can observe a homogeneous high perfusion region centered on HG. The contour of this bilateral region of high perfusion in AC was outlined in black on all other maps.

the presumed location of the PAC. The correspondence is appreciable by contouring the high perfusion region and superimposing such contour to the tonotopy (Figure 3.3A): the perfusion-based contour "segments" the V-shaped gradient of high-low-high frequency cutting through the low frequency region.

The overlap between this high perfusion region and the anatomically defined PAC was 72.15% (see Supporting Information, Figure 3.8).

3.4.4 Voice selective regions

Figure 3.4 shows the unthresholded voice selective maps in the CBA surface space obtained from CBF and BOLD time courses using RFX GLM and computing the contrast vocal sounds $>$ (other natural sounds + tones)/2. The overlaid black contours and the Supporting Information Figure 3.9A show the voice selective regions defined as those regions showing significantly higher activation to vocal sounds compared to other natural sounds and tones. BOLD signal voice selective regions presented several peaks of voice sensitivity, and in particular, on mid STS (lateral to HG), posterior STS and STG, and anterior STS for the left hemisphere only. Outside the temporal lobe, a significant cluster was detected bilaterally on the inferior frontal gyrus by both CBF and BOLD voice sensitive

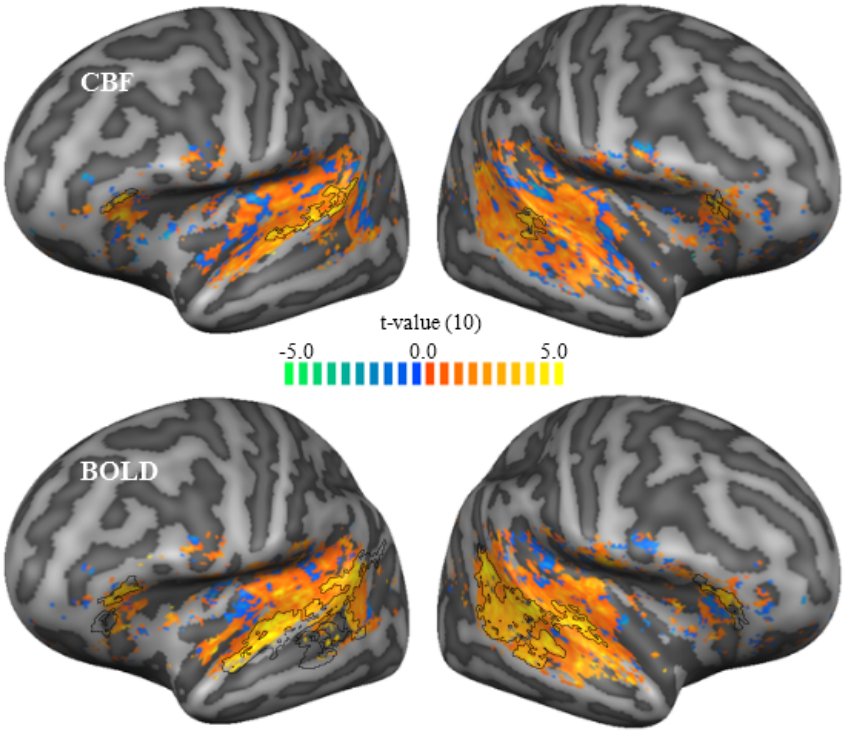

FIGURE 3.4: Unthresholded voice selective maps. Unthresholded CBF and BOLD voice selective maps (top and bottom row, respectively) were obtained using an RFX multi-subject GLM and computing the contrast vocal sounds > (other natural sounds + tones)/2 for the vertices which showed stimulus-induced activity for both CBF and BOLD signal (uncorrected *t*-value>2; contrast: all voice localizer sounds > baseline). The black contours indicate the voice selective regions obtained in the statistically thresholded maps computed in the Supporting Information (see Supporting Information Figure 3.9).

mapping (see Supporting Information Figures 3.10 and 3.11). This configuration is in agreement with previous studies, although an additional cluster on anterior STS in the right hemisphere is sometimes found (Belin et al., 2000; Moerel et al., 2012; Bonte et al., 2013, 2014; Pernet et al., 2015). In contrast, the extension of CBF voice selective regions was very limited, probably due to the lower CNR as suggested by below-threshold effects (see Supporting Information Figure 3.9B and Section 3.7.7 *Voice selective regions for a large range of initial vertex-level threshold* in the Supporting Information). To overcome the limitation of low SNR for CBF, the agreement between CBF and BOLD signal voice selectivity was assessed

calculating the Pearson's correlation between the two corresponding unthresholded maps. A correlation of 0.3815 ($p < 0.001$, two-tailed t-test) was found. Finally, good agreement was found between the BOLD signal voice selective map obtained in this study with that obtained by Pernet et al. (2015) using standard BOLD fMRI (Pearson's correlation of 0.4810; $p < 0.001$, two-tailed t-test).

3.4.5 Vein masks from susceptibility weighted images

We found no significant difference in the fraction of vein voxels when considering the whole ROI of BOLD versus CBF active voxels. However, considering the active voxels non-overlapping between the BOLD and the CBF active ROIs, we found a significantly higher fraction of vein voxels for BOLD exclusively active voxels versus CBF exclusively active voxels ($35\%\pm2\%$ and $27\%\pm2\%$, respectively; t-test(11) = 3.6114; $p = 0.0041$).

Hypothesizing a relationship between vein biasing and BOLD-CBF tonotopy mismatch, we calculated the correlation between BOLD and CBF tonotopy splitting the voxels according to the vein mask labelling. Vein voxels had a BOLD-CBF tonotopy correlation of 0.142 ± 0.039 and non-vein voxels of 0.163 ± 0.046 resulting in a non-significant difference (t-test(11) = 0.5169; $p = 0.3077$).

3.5 Discussion

The present study investigated the tonotopic organization of the human auditory cortex using ASL fMRI. In contrast to standard BOLD fMRI, ASL allows to simultaneously measuring CBF and BOLD signal. CBF has the advantage of being quantitative and physiologically meaningful, having higher spatial specificity and reproducibility, albeit with lower SNR compared to the BOLD signal. ASL has been previously employed to map the topography of a sensory system, namely retinotopy in visual cortex (Cavusoglu et al., 2012), while the current study, to the best of our knowledge, is the first to employ the ASL technique to perform tonotopic mapping in the auditory cortex.

As expected, the passive listening of the stimulus tones activated the auditory cortex bilaterally in a wide range of areas, such as HG, PT, PP, STG and STS. The extent of the activated areas was significantly larger when estimated from the BOLD signal than from CBF. Such difference was primarily most likely due to the different inherent SNR of the two contrasts, which is (for typical 3T human

imaging parameters) 3-5 times lower for CBF than for BOLD signal (Cavusoglu et al., 2012). In the current study, CNR of the BOLD signal was approximately twice as large as the CNR of CBF. These numbers are in agreement with the number of voxels detected as significantly active in this study. Despite the lower SNR, we observed a lower coefficient of variation for CBF percent signal change compared to BOLD, implying a higher reproducibility of CBF values, in agreement with previous studies (Tjandra et al., 2005; Leontiev and Buxton, 2007).

We demonstrated the feasibility of tonotopic mapping using CBF signal measured with ASL technique, specifically pCASL at 3T. The CBF tonotopy clearly showed a main V-shaped gradient of high frequencies around a low frequency region centered on HG and additional gradients in surrounding areas. The overall pattern of the tonotopic map was similar across the two hemispheres. The CBF tonotopy was in good agreement with the BOLD tonotopy obtained by the BOLD time course extracted from the same ASL signal. The BOLD tonotopy obtained from the ASL sequence, in turn, agreed very well with those of previous studies employing GE-EPI BOLD sequences (e.g., Formisano et al., 2003; Humphries et al., 2010; Da Costa et al., 2011; Moerel et al., 2012; Saenz and Langers, 2014). Although acquired simultaneously and therefore susceptible to the same correlated artifacts (e.g., physical noise, motion, ...), ideally, the BOLD signal and CBF represent physically independent modulation of the ASL signal. That is, the presence of tonotopy in both the BOLD signal and CBF provides reciprocal validation of the utility of both parameters for probing the human AC.

Despite the good correspondence between the spatial locations of the gradients between BOLD and CBF tonotopy, we observed a less steep gradient between the two extremes of the frequency scale using CBF, resulting in smaller areas activated preferentially by the lowest or highest frequencies. We attribute this discrepancy to the inherently lower SNR of CBF resulting in noisier single-subject maps and therefore favoring intermediate frequency range due to the averaging of best frequency values involved in the calculation of group maps.

To focus our analysis on the auditory core, we defined a mask including HG and the areas immediately surrounding HG anteriorly and posteriorly (see Figure 3.2A). The mask was anatomically defined in the CBA group space on the basis of previous literature and current best practice (Humphries et al., 2010; Langers and van Dijk, 2012; Moerel et al., 2014). The rationale of focusing on the auditory core is that this is the area of AC that, despite the debate about the orientation of PAC, is most consistently described and reliably interpreted across different

studies at different field strength, using different stimuli, and at single-subject and group level (Moerel et al., 2014). Interestingly, restricting the analysis to the anatomically defined PAC increased the spatial correlation between BOLD and CBF group tonotopic maps. This result confirms that human tonotopic maps — as obtained with tones — are more reliable and more consistent across subjects in PAC than in the whole AC. In future studies, it would be interesting to see whether combining CBF-based tonotopic mapping and natural sounds, which engage the whole auditory cortex in an ecologically valid manner (Moerel et al., 2012), increases consistency also outside PAC.

Additionally to the functional ASL runs, an M_0 image was acquired to allow the quantification of brain perfusion. On the basis of the quantification formula proposed by the ASL white paper (Equation 3.1; Alsop et al., 2015), we estimated the quantitative baseline CBF voxel-by-voxel using the full ASL model to obtain a perfusion estimate unbiased by the auditory activation due to the tone presentation. In the auditory cortex, we observed a region characterized by higher baseline perfusion values in the location of HG and immediate surroundings for both hemispheres. This finding is in agreement with the values of regional CBF reported by (Chen et al., 2011, cfr. Table 2 "Transverse temporal - young adults" and Figure 2), although in the cited paper no specific comment was done on this regard. At least two alternative explanations of the observed higher CBF in the presumed auditory core are possible: One possible cause is the noise of the MR gradients during image acquisition. Alternatively, relatively higher CBF is due to higher vascularization in the auditory core and thus independent of the MRI acquisition. In other words, even though the MR gradient noise is a stimulus for the auditory cortex, the spatial distribution of relative CBF can be preserved under the MRI conditions. Regardless of the underlying cause, we suggest that such high localized perfusion area detected bilaterally in the auditory cortex identifies the primary auditory core (the homologues of monkey areas A1 and R). This interpretation is supported by previous findings that primary (visual, auditory, and somatosensory) areas have higher vascular density and steady-state metabolic demands than secondary areas (Weber et al., 2008), and by the correspondence between the location of the high perfusion area and that of the main high-low-high frequency gradients of both BOLD and CBF tonotopies. Interestingly, the anterior border of the high perfusion area cuts through the low preferred frequency area of the gradient. This offers a possible distinction between primary

and non-primary auditory regions otherwise not possible on the basis of the tonotopic information alone. In conclusion, independently from its cause, we suggest that the observed higher perfusion is spatially restricted to early auditory areas, thus, allowing the parcellation of the auditory cortex.

We investigated the tissue specificity of BOLD and CBF signals and found a significantly higher number of vein voxels among BOLD signal active voxels, compared to when using CBF. Vein voxels were defined using vein masks created from SWI images optimized to enhance venous vessels from the surrounding tissue. Our results are in agreement with previous studies reporting a venous bias of the BOLD signal whilst a higher specificity to the capillary beds for CBF signal (Aguirre et al., 2002; Wang et al., 2003; Tjandra et al., 2005). We hypothesized that such venous bias of the BOLD signal could explain the observed BOLD signal-CBF tonotopy mismatches, but no significance difference was found between the BOLD signal-CBF tonotopy correlations of vein versus non-vein voxels. Further investigations are needed to shed light on the origin(s) of the observed mismatches between BOLD signal- and CBF-based tonotopy possibly using higher spatial resolution reducing partial volume effects between tissue, veins and CSF, in particular outside of the PAC.

To further characterize the human auditory cortex, we investigated a higher order functional property such as voice sensitivity. Voice selective regions were investigated by contrasting responses to vocal sounds versus those to other natural sounds and tones as measured by CBF and BOLD signal computed from the ASL signal of the voice localizer run. BOLD signal defined voice selective regions were mainly located on STG and STS and presented five peaks of voice sensitivity, namely posterior and mid STS for both hemispheres and anterior STS for the left hemisphere in good agreement with previous studies (Belin et al., 2000; Moerel et al., 2012; Bonte et al., 2013, 2014; Pernet et al., 2015). CBF defined voice selective regions, although showing a more limited extent, successfully detected three peaks corresponding to the bilateral posterior and the left mid STS clusters. The correlation analysis between the unthresholded BOLD- and CBF-based voice selective maps showed their relatively good agreement and further supported the hypothesis that differences in extent and number of detected peaks was most likely due to the different CNR of the BOLD and CBF signal.

3.5.1 Limitations and benefits of tonotopy using ASL

The most stringent limitation of using CBF signal is its low SNR (compared to the standard BOLD signal). In this study, we assessed the CNR as a measure of the functional sensitivity of the data and we found a CNR 1.6 times lower for CBF than BOLD signal. Thus, differences between auditory processing as detected using CBF and BOLD signal can either be attributed to the differences in the biophysical origins of both signals or to differences in their CNRs. In future studies, an adequately larger number of trials could be used to overcome the ASL CNR penalty. Moreover, the labeling duration and PLD used in this study were shorter than those recommended by (Alsop et al., 2015). Using a pCASL sequence with longer labeling duration and PLD might have resulted in higher SNR of the baseline CBF. Note, however, even though some of the quantitative results on image SNR and tSNR of CBF are affected by the choice of ASL parameters, the results on CBF tonotopic maps and their comparison with BOLD signal tonotopic maps are qualitatively insensitive for a wide range of these parameters.

Another limitation of using ASL techniques is the need of acquiring tag-control pairs of images, which results in an effective temporal resolution lower than the nominal TR. Moreover, the TR itself cannot be as short as in BOLD imaging because of the post-labeling delay to allow the blood to reach the imaging slab. Such transit time constitutes a time constraint of the order of $\sim$700-2800 ms depending of the region of interest (Mildner et al., 2014), which, however, enables presenting the auditory stimuli within the silent period of the delay.

On the other hand, CBF offers some important advantages such as quantification, physiological unit of measure, reproducibility and spatial specificity. Moreover, our results show that the baseline perfusion signal offers additional information to characterize AC. Most importantly, delineating the primary auditory core on the basis of the perfusion baseline map alone provides complementary and independent information to anatomical landmarks or myelin delineations. ASL perfusion baseline measurements can be performed without sound presentation and with a run duration of 3-10 minutes (depending on the spatial resolution), therefore in a much shorter acquisition time than usual tonotopy protocols. FMRI studies interested in PAC localization, but not in tonotopic information per se, could therefore greatly benefit from perfusion baseline PAC delineation as they could invest the spared time in the effect/task of interest. In addition, differences in baseline perfusion between populations (e.g., healthy subjects vs

Tinnitus patients) may be detected and be meaningful in characterizing the state of the auditory processing in these populations. Furthermore, venous biases potentially confounding BOLD signal maps (such as detected in V4 in the visual cortex; see Winawer et al., 2010) may be absent in CBF maps, which therefore may yield a more faithful representation of the underlying neuronal functional architecture.

3.6 Conclusions

In this study, we demonstrated the feasibility of tonotopy and voice area mapping in human auditory cortex using CBF obtained with an ASL MRI sequence. We described the limitations and benefits of this approach compared to standard BOLD fMRI: CBF is characterized by a lower CNR and temporal resolution, but is a quantifiable physiological measure, has higher reproducibility, higher spatial specificity, and ASL sequences allow the simultaneous acquisition of CBF and BOLD signal and sound presentation during the silent PLD. Interpreting the perfusion baseline map and tonotopy together, we propose quantitative baseline perfusion as a novel marker to identify the primary auditory cortex.

Bibliography

G. K. Aguirre, J. A. Detre, E. Zarahn, and D. C. Alsop. Experimental design and the relative sensitivity of BOLD and perfusion fMRI. *Neuroimage*, 15(3):488–500, 2002.

D. C. Alsop, J. A. Detre, X. Golay, M. Gunther, J. Hendrikse, L. Hernandez-Garcia, H. Lu, B. J. MacIntosh, L. M. Parkes, M. Smits, M. J. van Osch, D. J. Wang, E. C. Wong, and G. Zaharchuk. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. *Magn Reson Med*, 73(1):102–16, 2015.

B. Barton, J. H. Venezia, K. Saberi, G. Hickok, and A. A. Brewer. Orthogonal acoustic dimensions define auditory field maps in human cortex. *Proc Natl Acad Sci U S A*, 109(50):20738–43, 2012.

S. Baumann, C. I. Petkov, and T. D. Griffiths. A unified framework for the organization of the primate auditory cortex. *Front Syst Neurosci*, 7:11, 2013.

P. Belin, R. J. Zatorre, P. Lafaille, P. Ahad, and B. Pike. Voice-selective areas in human auditory cortex. *Nature*, 403(6767):309–12, 2000.

N. A. Bock, A. Kocharyan, J. V. Liu, and A. C. Silva. Visualizing the entire cortical myelination pattern in marmosets with magnetic resonance imaging. *J Neurosci Methods*, 185(1):15–22, 2009.

M. Bonte, M. A. Frost, S. Rutten, A. Ley, E. Formisano, and R. Goebel. Development from childhood to adulthood increases morphological and functional inter-individual variability in the right superior temporal cortex. *Neuroimage*, 83:739–50, 2013.

M. Bonte, L. Hausfeld, W. Scharke, G. Valente, and E. Formisano. Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *J Neurosci*, 34(13):4548–57, 2014.

M. Cavusoglu, J. Pfeuffer, K. Ugurbil, and K. Uludag. Comparison of pulsed arterial spin labeling encoding schemes and absolute perfusion quantification. *Magn Reson Imaging*, 27(8):1039–45, 2009.

M. Cavusoglu, A. Bartels, B. Yesilyurt, and K. Uludag. Retinotopic maps and hemodynamic delays in the human visual cortex measured using arterial spin labeling. *Neuroimage*, 59(4):4044–54, 2012.

J. J. Chen, H. D. Rosas, and D. H. Salat. Age-associated reductions in cerebral blood flow are independent from regional atrophy. *Neuroimage*, 55(2):468–78, 2011.

J. Cohen-Adad, J. R. Polimeni, K. G. Helmer, T. Benner, J. A. McNab, L. L. Wald, B. R. Rosen, and C. Mainero. T2* mapping and B0 orientation-dependence at 7 T reveal cyto- and myeloarchitecture organization of the human cortex. *Neuroimage*, 60(2):1006–14, 2012.

S. Da Costa, W. van der Zwaag, J. P. Marques, R. S. Frackowiak, S. Clarke, and M. Saenz. Human primary auditory cortex follows the shape of Heschl's gyrus. *J Neurosci*, 31(40):14067–75, 2011.

W. Dai, D. Garcia, C. de Bazelaire, and D. C. Alsop. Continuous flow-driven inversion for arterial spin labeling using pulsed radio frequency and gradient fields. *Magn Reson Med*, 60(6):1488–97, 2008.

F. De Martino, M. Moerel, P. F. van de Moortele, K. Ugurbil, R. Goebel, E. Yacoub, and E. Formisano. Spatial organization of frequency preference and selectivity in the human inferior colliculus. *Nat Commun*, 4:1386, 2013.

F. De Martino, M. Moerel, J. Xu, P. F. van de Moortele, K. Ugurbil, R. Goebel, E. Yacoub, and E. Formisano. High-resolution mapping of myeloarchitecture in vivo: Localization of auditory areas in the human brain. *Cereb Cortex*, 25(10): 3394–405, 2015.

F. Dick, A. T. Tierney, A. Lutti, O. Josephs, M. I. Sereno, and N. Weiskopf. In vivo functional and myeloarchitectonic mapping of human primary auditory areas. *J Neurosci*, 32(46):16095–105, 2012.

M. J. Donahue, C. C. Faraco, M. K. Strother, M. A. Chappell, S. Rane, L. M. Dethrage, J. Hendrikse, and J. C. Siero. Bolus arrival time and cerebral blood flow responses to hypercarbia. *J Cereb Blood Flow Metab*, 34(7):1243–52, 2014.

B. Fischl, M. I. Sereno, R. B. Tootell, and A. M. Dale. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp*, 8 (4):272–84, 1999.

E. Formisano, D. S. Kim, F. Di Salle, P. F. van de Moortele, K. Ugurbil, and R. Goebel. Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, 40(4):859–69, 2003.

A. Gardumi, D. Ivanov, L. Hausfeld, G. Valente, E. Formisano, and K. Uludag. The effect of spatial resolution on decoding accuracy in fMRI multivariate pattern analysis. *Neuroimage*, 132:32–42, 2016.

S. Geyer, M. Weiss, K. Reimann, G. Lohmann, and R. Turner. Microstructural parcellation of the human cerebral cortex - From Brodmann's post-mortem map to in vivo mapping with high-field magnetic resonance imaging. *Front Hum Neurosci*, 5:19, 2011.

M. F. Glasser and D. C. Van Essen. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. *J Neurosci*, 31(32): 11597–616, 2011.

R. Goebel, F. Esposito, and E. Formisano. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp*, 27(5):392–401, 2006.

E. M. Haacke, Y. Xu, Y. C. Cheng, and J. R. Reichenbach. Susceptibility weighted imaging (SWI). *Magn Reson Med*, 52(3):612–8, 2004.

J. Harmer, R. M. Sanchez-Panchuelo, R. Bowtell, and S. T. Francis. Spatial location and strength of BOLD activation in high-spatial-resolution fMRI of the motor cortex: a comparison of spin echo and gradient echo fMRI at 7 T. *NMR Biomed*, 25(5):717–25, 2012.

M. Herdener, F. Esposito, K. Scheffler, P. Schneider, N. K. Logothetis, K. Uludag, and C. Kayser. Spatial representations of temporal and spectral sound cues in human auditory cortex. *Cortex*, 49(10):2822–33, 2013.

L. Hernandez-Garcia, H. Jahanian, and D. B. Rowe. Quantitative analysis of arterial spin labeling fMRI data using a general linear model. *Magn Reson Imaging*, 28(7):919–27, 2010.

C. Humphries, E. Liebenthal, and J. R. Binder. Tonotopic organization of human auditory cortex. *Neuroimage*, 50(3):1202–11, 2010.

J. J. Kim, B. Crespo-Facorro, N. C. Andreasen, D. S. O'Leary, B. Zhang, G. Harris, and V. A. Magnotta. An MRI-based parcellation method for the temporal lobe. *Neuroimage*, 11(4):271–88, 2000.

D. R. Langers and P. van Dijk. Mapping the tonotopic organization in human auditory cortex with minimally salient acoustic stimulation. *Cereb Cortex*, 22 (9):2024–38, 2012.

D. R. Langers, W. H. Backes, and P. van Dijk. Representation of lateralization and tonotopy in primary versus secondary human auditory cortex. *Neuroimage*, 34 (1):264–73, 2007.

O. Leontiev and R. B. Buxton. Reproducibility of BOLD, perfusion, and CMRO2 measurements with calibrated-BOLD fMRI. *Neuroimage*, 35(1):175–84, 2007.

T. T. Liu and E. C. Wong. A signal processing model for arterial spin labeling functional MRI. *Neuroimage*, 24(1):207–15, 2005.

N. K. Logothetis. What we can do and what we cannot do with fMRI. *Nature*, 453 (7197):869–78, 2008.

H. Lu, C. Clingman, X. Golay, and P. C. van Zijl. Determining the longitudinal relaxation time (T1) of blood at 3.0 Tesla. *Magn Reson Med*, 52(3):679–82, 2004.

M. M. Merzenich and J. F. Brugge. Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Res*, 50(2):275–96, 1973.

M. M. Merzenich, P. L. Knight, and G. L. Roth. Cochleotopic organization of primary auditory cortex in the cat. *Brain Res*, 63:343–6, 1973.

M. Mezue, A. R. Segerdahl, T. W. Okell, M. A. Chappell, M. E. Kelly, and I. Tracey. Optimization and reliability of multiple postlabeling delay pseudo-continuous arterial spin labeling during rest and stimulus-induced functional task activation. *J Cereb Blood Flow Metab*, 34(12):1919–27, 2014.

T. Mildner, K. Muller, S. Hetzer, R. Trampel, W. Driesel, and H. E. Moller. Mapping of arterial transit time by intravascular signal selection. *NMR Biomed*, 27 (5):594–609, 2014.

M. Moerel, F. De Martino, and E. Formisano. Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. *J Neurosci*, 32(41):14205–16, 2012.

M. Moerel, F. De Martino, R. Santoro, K. Ugurbil, R. Goebel, E. Yacoub, and E. Formisano. Processing of natural sounds: characterization of multipeak spectral tuning in human auditory cortex. *J Neurosci*, 33(29):11888–98, 2013.

M. Moerel, F. De Martino, and E. Formisano. An anatomical and functional topography of human auditory cortical areas. *Front Neurosci*, 8:225, 2014.

J. P. Mugler and J. R. Brookeman. Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *Magn Reson Med*, 15(1):152–7, 1990.

J. A. Mumford, L. Hernandez-Garcia, G. R. Lee, and T. E. Nichols. Estimation efficiency and statistical power in arterial spin labeling fMRI. *Neuroimage*, 33 (1):103–14, 2006.

C. R. Pernet, P. McAleer, M. Latinus, K. J. Gorgolewski, I. Charest, P. E. Bestelmeyer, R. H. Watson, D. Fleming, F. Crabbe, M. Valdes-Sosa, and P. Belin. The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage*, 119:164–74, 2015.

A. Rauscher, J. Sedlacik, A. Deistung, H. J. Mentzel, and J. R. Reichenbach. Susceptibility weighted imaging: data acquisition, image reconstruction and clinical applications. *Z Med Phys*, 16(4):240–50, 2006.

J. R. Reichenbach, R. Venkatesan, D. J. Schillinger, D. K. Kido, and E. M. Haacke. Small vessels in the human brain: MR venography with deoxyhemoglobin as an intrinsic contrast agent. *Radiology*, 204(1):272–7, 1997.

M. Saenz and D. R. Langers. Tonotopic mapping of human auditory cortex. *Hear Res*, 307:42–52, 2014.

M. I. Sereno, A. Lutti, N. Weiskopf, and F. Dick. Mapping the human cortical surface by combining quantitative T1 with retinotopy. *Cereb Cortex*, 23(9):2261–8, 2013.

I. S. Sigalovsky, B. Fischl, and J. R. Melcher. Mapping an intrinsic MR property of gray matter in auditory cortex of living humans: a possible marker for primary cortex and hemispheric differences. *Neuroimage*, 32(4):1524–37, 2006.

T. M. Talavage, M. I. Sereno, J. R. Melcher, P. J. Ledden, B. R. Rosen, and A. M. Dale. Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *J Neurophysiol*, 91(3):1282–96, 2004.

T. Tjandra, J. C. Brooks, P. Figueiredo, R. Wise, P. M. Matthews, and I. Tracey. Quantitative assessment of the reproducibility of functional activation measured with BOLD and MR perfusion imaging: implications for clinical trial design. *Neuroimage*, 27(2):393–401, 2005.

K. Ugurbil, L. Toth, and D. S. Kim. How accurate is magnetic resonance imaging of brain function? *Trends Neurosci*, 26(2):108–14, 2003.

K. Uludag, B. Muller-Bierl, and K. Ugurbil. An integrative model for neuronal activity-induced signal changes for gradient and spin echo functional imaging. *Neuroimage*, 48(1):150–65, 2009.

J. Wang, G. K. Aguirre, D. Y. Kimberg, A. C. Roc, L. Li, and J. A. Detre. Arterial spin labeling perfusion fMRI with very low task frequency. *Magn Reson Med*, 49(5):796–802, 2003.

B. Weber, A. L. Keller, J. Reichold, and N. K. Logothetis. The microvascular system of the striate and extrastriate visual cortex of the macaque. *Cereb Cortex*, 18 (10):2318–30, 2008.

M. Welvaert and Y. Rosseel. On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data. *PLoS One*, 8(11):e77089, 2013.

J. Winawer, H. Horiguchi, R. A. Sayres, K. Amano, and B. A. Wandell. Mapping hV4 and ventral occipital cortex: the venous eclipse. *J Vis*, 10(5):1, 2010.

D. L. Woods, T. J. Herron, A. D. Cate, E. W. Yund, G. C. Stecker, T. Rinne, and X. Kang. Functional properties of human auditory cortical fields. *Front Syst Neurosci*, 4:155, 2010.

X. Zhang, E. T. Petersen, E. Ghariq, J. B. De Vis, A. G. Webb, W. M. Teeuwisse, J. Hendrikse, and M. J. van Osch. In vivo blood T1 measurements at 1.5 T, 3 T, and 7 T. *Magn Reson Med*, 70(4):1082–6, 2013.

3.7 Supporting Information

3.7.1 Vein masks

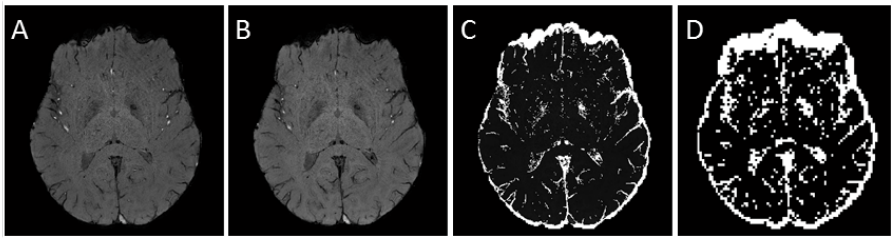

FIGURE 3.5: Vein mask creation for a representative subject. (A) Original SW image. (B) mIP-SW image. (C) Thresholded and binarised mIP-SW image. (D) Final vein mask obtained after coregistration in the native anatomical space of the subject and downsampling to 2.0 mm isotropic voxel size. Note that CSF is included in this mask, which, however, is not expected to affect the results obtained from the cortical surface.

3.7.2 Power spectrum of 3T pCASL with 2D single-shot readout

The noise produced by the 3T scanner during the pCASL acquisition with 2D single-shot EPI readout was measured using an optical microphone positioned inside the MR coil at the approximate location of the head of the participants. In Figure 3.1 of the manuscript, Panel A shows the sound waveform for one TR of the used sequence and Panel B the frequency spectrum for the tagging module and the EPI-train readout.

3.7.3 Activation in auditory cortex

As reported in the main manuscript, the boxplots in Figure 3.6 show the distributions across the subjects of the number of active voxels for BOLD signal and CBF.

3.7.4 CBF quantification

Baseline perfusion was quantified according to Equation 3.1 (Alsop et al., 2015) and the resulting quantitative perfusion maps are shown in Figure 3.7 as single-subject multi-slice view.

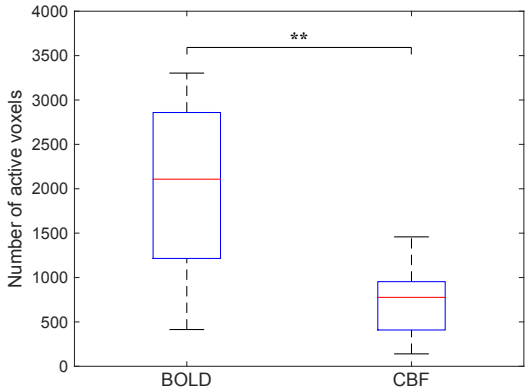

FIGURE 3.6: Number of active voxels for BOLD and CBF signal. Boxplot describing the distribution of active voxels across the subjects ("active" here means surviving a statistical threshold of $Q(FDR) < 0.05$ (for the contrast: sound vs baseline) and "**" indicates a significant difference of $p < 0.001$ (two-tailed t-test with $N = 12$ subjects)).

3.7.5 Anatomically defined and baseline CBF based PAC

Figure 3.8 shows in pink the anatomically defined primary auditory mask (same as in Figure 3.2A of the main text) and, overlaid on it, the black contour of the high perfusion region identified using the baseline CBF measure.

3.7.6 Voice selective regions

In the manuscript, unthresholded CBF and BOLD statistical maps were used to compute their spatial correlation as measure of their agreement. Here, we follow more standard processing and threshold them in order to obtain voice selective regions defined as those regions showing significantly higher activation to vocal sounds compared to other natural sounds and tones. To this aim, the statistics obtained from the RFX GLM analysis were contrasted using vocal sounds $>$ (other natural sounds + tones)/2 and the resulting maps were corrected for multiple comparisons using a cluster size thresholding approach (Forman et al., 1995; Goebel et al., 2006; Hagler et al., 2006). The initial vertex-level threshold was set to $p < 0.05$ (t-value(10) > 2.22) and a surface-based cluster size threshold was estimated with a false-positive rate (α) of 0.05: The estimated cluster size threshold for CBF was 20 mm^2 for both left and right hemispheres (LH and RH), while for

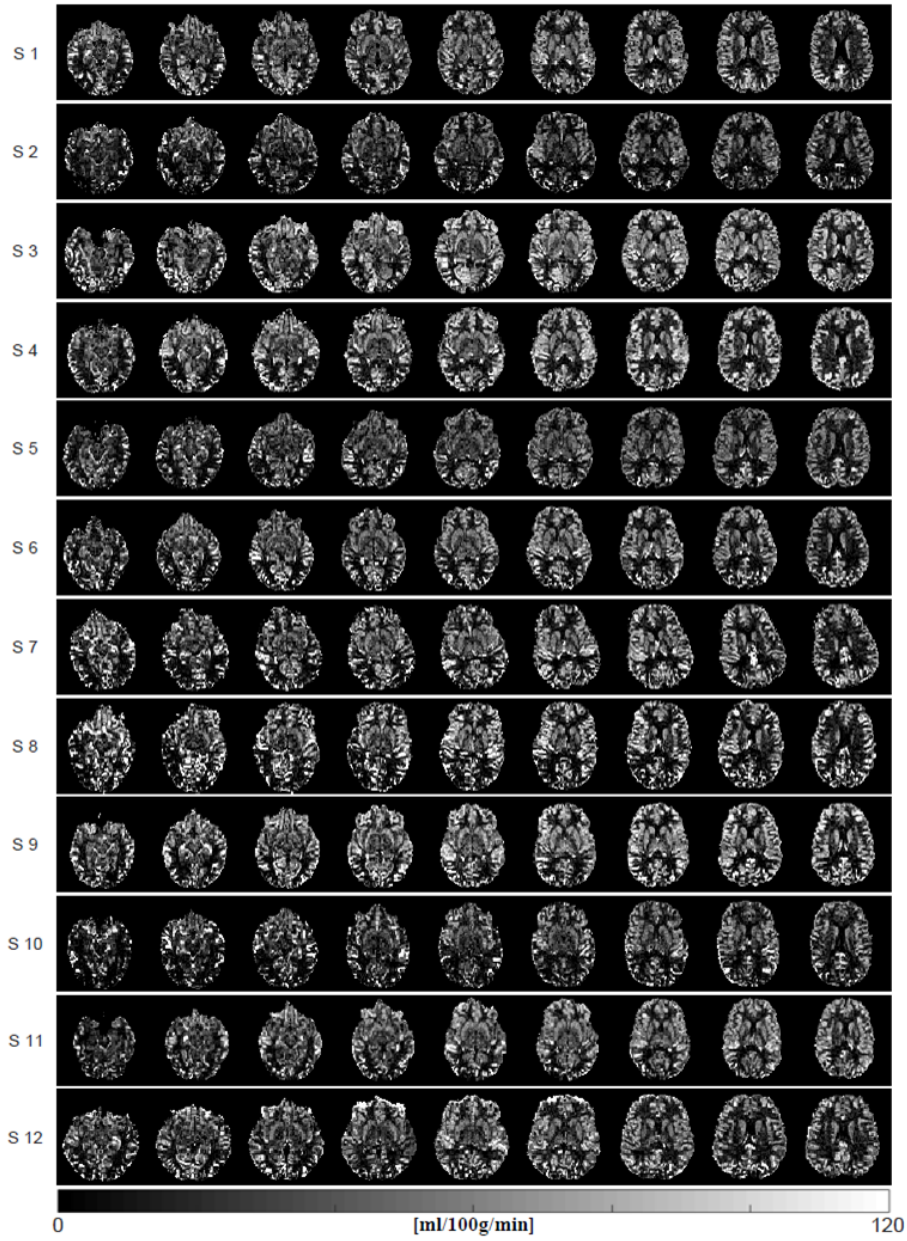

FIGURE 3.7: Single-subject quantitative perfusion maps. Each row shows a single-subject multi-slice view of the quantitative perfusion maps. All subjects show clear high gray matter perfusion and low white matter perfusion.

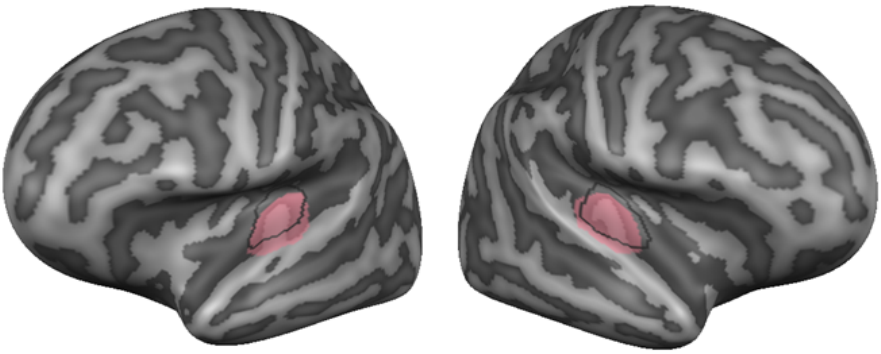

FIGURE 3.8: Comparison of the baseline CBF based definition of PAC (solid black line) and its anatomical definition (pink region).

the BOLD signal it was 48 and 49 mm^2 for LH and RH, respectively. Given the arbitrary choice of the initial vertex-level threshold and its large influence on the resulting maps, we additionally explored the resulting voice selectivity maps for a large range of initial vertex-level thresholds (for details, please see following section in the Supporting Information).

Figure 3.9A shows the group voice selective maps in the CBA surface space, masked with the anatomically defined temporal cortex mask (see blue mask in Figure 3.2A). Voice selective regions are mainly located on STG and STS in both hemispheres. BOLD signal voice selective regions presented several peaks of voice sensitivity, and in particular, on mid STS (lateral to HG), posterior STS and STG, and anterior STS for the left hemisphere only. In contrast, the extension of CBF voice selective regions was very limited, probably due to the lower CNR as suggested by below-threshold effects (see Figure 3.9B and following section in the Supporting Information). Nevertheless, when detected, the location of the CBF-based voice selective regions was in agreement with the peak of the BOLD-based voice selective regions. Specifically, the detected locations were: the mid STS peak in the left hemisphere and the posterior STS/STG in both hemispheres.

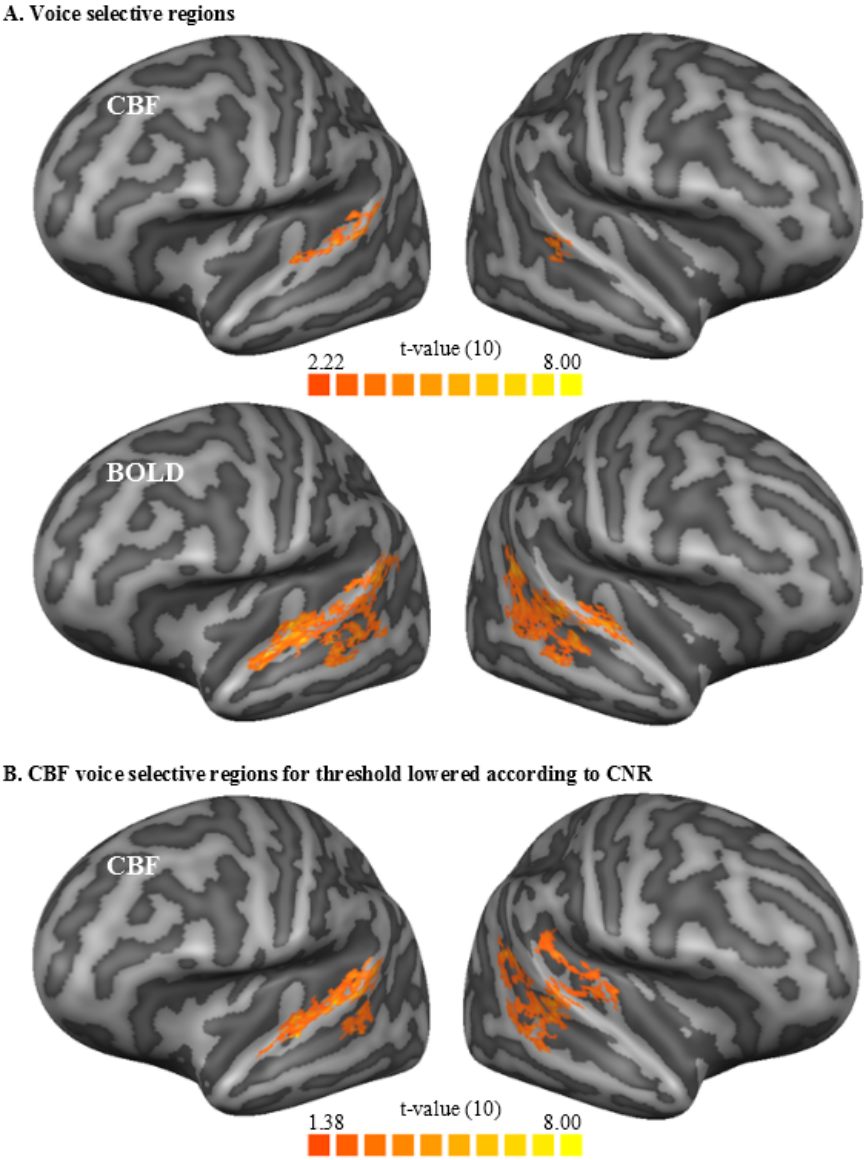

FIGURE 3.9: Group voice selective regions as estimated by the RFX multi-subject GLM. Panel A shows the CBF and BOLD voice selective regions (top and bottom row, respectively) estimated using the same initial threshold of t-value(10) > 2.22. Panel B shows the CBF voice selective map resulting from an initial threshold lowered proportionally to the ratio between BOLD and CBF CNR (see following section in the Supporting Information for further details about this explanatory analysis). Maps were masked with the anatomically defined temporal cortex mask.

3.7.7 Voice selective regions for a large range of initial vertex-level threshold

In the previous section, voice selective regions were defined as those regions showing significantly higher activation to vocal sounds compared to other natural sounds and tones. The maps were corrected for multiple comparisons using a cluster size thresholding approach (Forman et al., 1995; Goebel et al., 2006; Hagler et al., 2006). Here, we explore a large range of initial vertex-level thresholds adequately chosen for CBF and BOLD signal CNR. In Figure 3.10 and 3.11, maps are reported without anatomical masking and, as reference, a dashed rectangle highlights the map corresponding to the standard threshold chosen in the previous section (i.e., $p < 0.05$ corresponding to t-value(10) > 2.22). Note that the case "$t > 1.38$ CBF map" corresponds to lowering the initial t-value thresholding proportionally to the ratio between BOLD and CBF CNR.

As expected, for both CBF and BOLD signal the extension of voice selective regions increased for lower thresholds and, vice versa, decreased for higher thresholds. Lowering the initial threshold for the CBF-based map yielded results more similar to the BOLD-based "reference (i.e., the $p < 0.05$, $t > 2.22$) map": for low thresholds all five voice sensitive clusters (i.e., anterior STG in the left hemisphere and mid and posterior STG in both hemispheres) were detected; however, some additional clusters on HG and posteriorly to it also survived this milder threshold although they were not present in the BOLD-based reference map and previous studies (Belin et al., 2000; Moerel et al., 2012; Bonte et al., 2013, 2014). Similarly, increasing the threshold for the BOLD-based map yielded results more similar to the CBF-based "reference (i.e., the $p < 0.05$, $t > 2.22$) map". Qualitatively, the "$t > 2.22$ CBF map" was similar to the "$t > 3.00$ BOLD map" for the left hemisphere and the "$t > 3.75$ BOLD map" for the right hemisphere; the "$t > 2.22$ BOLD map" was similar to the "$t > 1.20$ CBF map" for both hemispheres. This explorative analysis had the qualitative purpose to exemplify the influence of the different CNR between CBF and BOLD signal on the resulting voice selective regions. Importantly, these observations allowed us to attribute (main part of) the differences in the voice selective clusters detected by CBF and BOLD signal to their different CNR.

Voice selective maps are reported in Figure 3.10 and 3.11 without any anatomical mask, showing the presence of a bilateral extra-temporal cluster on the inferior frontal gyrus detected in agreement with (Pernet et al., 2015).

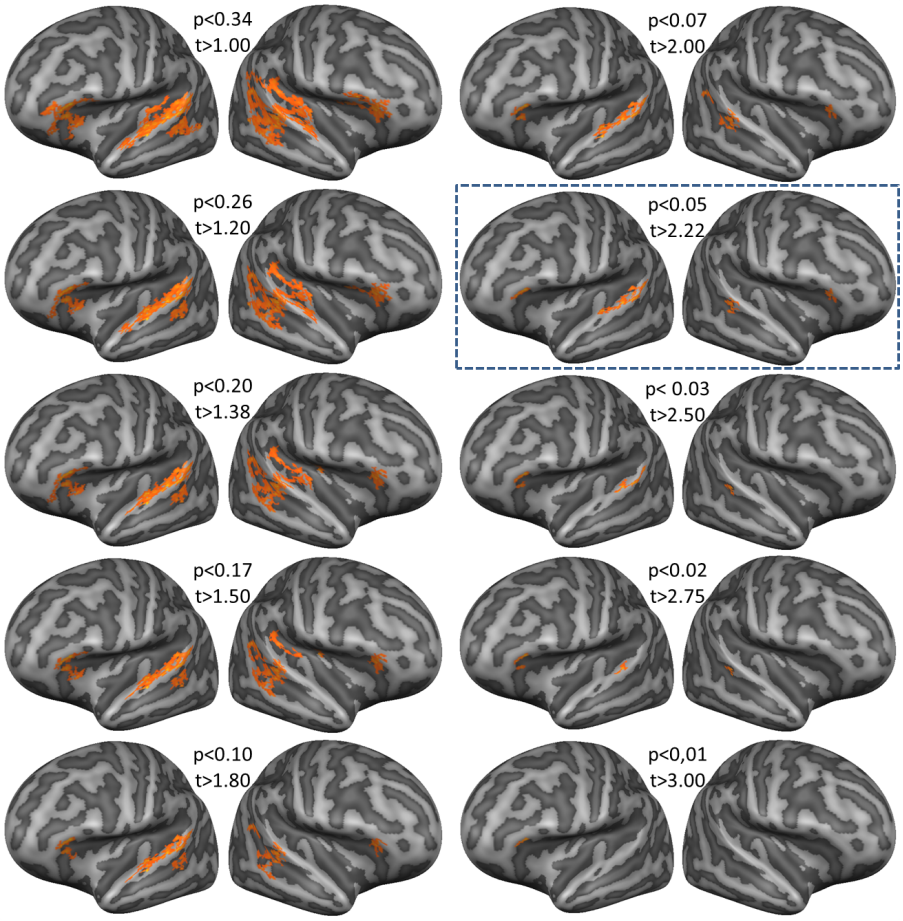

FIGURE 3.10: CBF voice selective regions for a large range of initial vertex-level threshold. The dashed rectangle indicates the standard threshold taken as reference ($p < 0.05, t > 2.22$).

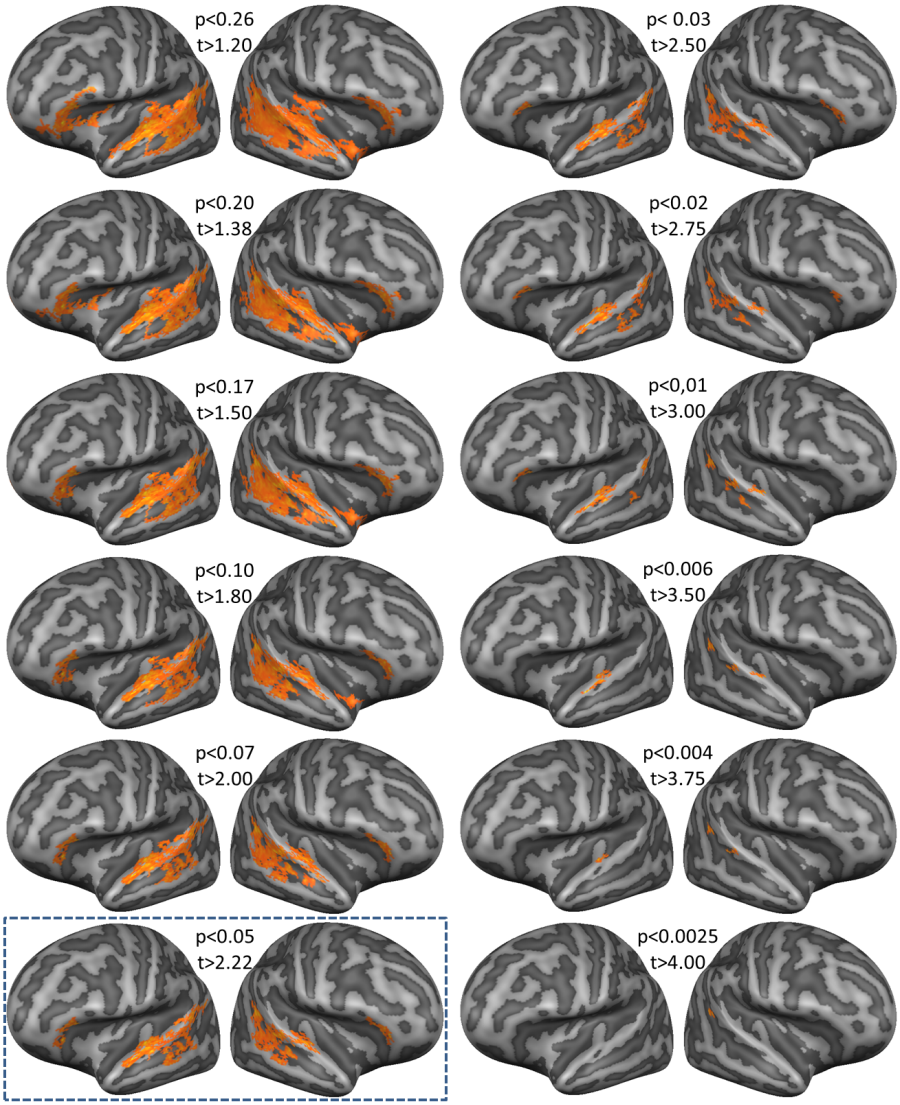

FIGURE 3.11: BOLD signal voice selective regions for a large range of initial vertex-level threshold. The dashed rectangle indicates the standard threshold taken as reference ($p < 0.05$, $t > 2.22$).

Bibliography

D. C. Alsop, J. A. Detre, X. Golay, M. Gunther, J. Hendrikse, L. Hernandez-Garcia, H. Lu, B. J. MacIntosh, L. M. Parkes, M. Smits, M. J. van Osch, D. J. Wang, E. C. Wong, and G. Zaharchuk. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. *Magn Reson Med*, 73(1):spcone, 2015.

P. Belin, R. J. Zatorre, P. Lafaille, P. Ahad, and B. Pike. Voice-selective areas in human auditory cortex. *Nature*, 403(6767):309–12, 2000.

M. Bonte, M. A. Frost, S. Rutten, A. Ley, E. Formisano, and R. Goebel. Development from childhood to adulthood increases morphological and functional inter-individual variability in the right superior temporal cortex. *Neuroimage*, 83:739–50, 2013.

M. Bonte, L. Hausfeld, W. Scharke, G. Valente, and E. Formisano. Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *J Neurosci*, 34(13):4548–57, 2014.

S. D. Forman, J. D. Cohen, M. Fitzgerald, W. F. Eddy, M. A. Mintun, and D. C. Noll. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn Reson Med*, 33(5): 636–47, 1995.

R. Goebel, F. Esposito, and E. Formisano. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp*, 27(5):392–401, 2006.

Jr. Hagler, D. J., A. P. Saygin, and M. I. Sereno. Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *Neuroimage*, 33(4): 1093–103, 2006.

M. Moerel, F. De Martino, and E. Formisano. Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. *J Neurosci*, 32(41):14205–16, 2012.

C. R. Pernet, P. McAleer, M. Latinus, K. J. Gorgolewski, I. Charest, P. E. Bestelmeyer, R. H. Watson, D. Fleming, F. Crabbe, M. Valdes-Sosa, and P. Belin. The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage*, 119:164–74, 2015.

Chapter 4

Comparison of 3T and 7T ASL for concurrent CBF and BOLD imaging

Based on:

Ivanov*, D., Gardumi*, A., Haast, R., Pfeuffer, J., Poser, B.A., and Uludağ, K. (under review). *Comparison of 3 and 7 Tesla ASL techniques for concurrent functional perfusion and BOLD studies*.

*Authors contributed equally to the work

4.1 Abstract

Arterial spin labeling (ASL) is the primary non-invasive MRI approach to measure baseline cerebral blood flow (CBF) in healthy subjects and patients. ASL also allows concurrent functional BOLD signal and CBF measurements, but the latter typically suffers from low contrast-to-noise (CNR) ratio. Ultra-high field imaging significantly boosts BOLD signal CNR and has the potential to also improve CBF CNR. However, it is contested whether also CBF CNR benefits from increasing magnetic field strength, especially given that technical challenges exist related to field inhomogeneities and power deposition constraints. Recently, we presented an optimized PASL technique that utilizes tr-FOCI inversion pulses and dielectric pads to overcome the temporal resolution limitations of previous 7T ASL implementations (Ivanov et al., 2016a). The primary goal of this study was to compare its performance to that of 3T ASL approaches — both pulsed and pseudo-continuous ASL (PASL and pCASL, respectively) — concerning functional studies using simultaneous CBF and BOLD signal acquisition. To this aim, we investigated a wide range of parameters that can influence CBF and BOLD signal sensitivities: spatial resolution, labeling scheme, parallel imaging and echo time. We found that 7T ASL is superior in terms of CBF and BOLD temporal signal-to-noise ratio (SNR) and activation volume compared to all 3T ASL variants, in particular at high spatial resolution. Our results show that the advantages of 7T for ASL stem from increased image SNR and improved parallel imaging performance. The gray matter baseline CBF was in good agreement for all 3T ASL variants, but a significantly lower value was obtained at 7T, possibly due to reduced intra-vascular MRI signal. The labeling scheme utilized was also found to significantly influence the measured perfusion territory CBF. In conclusion, a single-echo accelerated 7T PASL is recommended for high spatial and temporal resolution CBF and BOLD imaging, while a 3T dual-echo pCASL approach without parallel imaging may be preferred for low (i.e., 3 mm isotropic and lower) resolution functional perfusion and BOLD applications.

4.2 Introduction

Cerebral blood flow (CBF) can be measured and quantified using arterial spin labeling (ASL) — a non-invasive MRI method, which employs magnetically labeled arterial blood water as an endogenous tracer (Detre et al., 1992; Williams et al., 1992). ASL time-series consist of pairs of label (or tag) and control (i.e., non-labeled) images, whose subtraction produces a signal proportional to the local tissue perfusion while their average represents the BOLD signal (Luh et al., 2000). The non-invasive nature of ASL and its ability to dynamically and quantitatively measure CBF make it an attractive approach for both neuroscience research and clinical applications (Detre et al., 1998; Detre and Wang, 2002). Baseline CBF has been used to study longitudinal intra-subject changes due to, for example, learning, experience and plasticity, and also to examine inter-subject differences in baseline brain physiology in healthy subjects and patients (Detre and Wang, 2002; Krieger et al., 2014). Functional ASL changes have been shown to be better localized to the site of neural activation than the BOLD signal and quantitatively more directly related to neuronal activity (Duong et al., 2001; Tjandra et al., 2005; Cavusoglu et al., 2012; Havlicek et al., 2015). In addition, the capability of ASL to concurrently measure CBF and BOLD signal has proven to be useful for investigating the brain's physiology in health and disease (Bulte et al., 2012; Buxton et al., 2014).

Nevertheless, compared to BOLD fMRI, CBF measurements using ASL present some drawbacks (Alsop et al., 2015), such as: 1) lower signal-to-noise ratio (SNR) of the perfusion-weighted signal due to the low microvascular density ($\sim$1–2 % of local tissue volume); 2) reduced temporal resolution due to the necessity of a post-labeling delay (PLD) to allow the labeled blood to reach the imaging slab and due to the need to acquire pairs of label and control images; 3) limited brain coverage due to T1 relaxation of the labeled blood; and 4) increased power deposition.

To overcome these limitations, different ASL approaches have been proposed. An exhaustive overview on how to tackle these limitations, together with protocol recommendations, is given in the ASL white paper (Alsop et al., 2015). In particular, the pseudo-continuous ASL (pCASL) method (Wu et al., 2007; Dai et al., 2008) allows prolonged labeling durations, thus raising SNR, while the use of background suppression improves SNR by reducing the physiological noise in the time-series (Ye et al., 2000) (although at the cost of BOLD sensitivity (Ghariq

et al., 2014)). In addition, performing ASL at ultra-high field (UHF; i.e., 7 Tesla and higher) promises to be advantageous (Gardener et al., 2009; Ivanov et al., 2016a). The image SNR increases with field strength (Norris, 2003; Pohmann et al., 2016), which, for instance, improves white matter (WM) perfusion measurements with ASL (Gardener and Jezzard, 2014). In addition, the increased longitudinal relaxation times at higher fields (Rooney et al., 2007; Wright et al., 2008) reduce label decay during PLD and image acquisition leading to higher perfusion SNR. These advantages allow acquiring larger brain coverage and/or using longer PLD to avoid vascular artifacts, which are due to incomplete transfer of the labeled blood from the arterial tree to the local tissue. The advantages of UHF for BOLD imaging, demonstrated in numerous studies (e.g., Uludag et al., 2009; van der Zwaag et al., 2009; Donahue et al., 2011, and references therein), render simultaneous CBF and BOLD imaging using ASL at 7T particularly attractive.

Despite the aforementioned gains, UHF ASL has not found widespread use in humans due to significant technical challenges. First, the spatial homogeneity and efficiency of the labeling are significantly degraded at UHF due to B_0- and B_1^+-inhomogeneities (Teeuwisse et al., 2010; Luh et al., 2013). Second, SAR constraints at 7T make not only spin-echo (SE) based approaches impractical, but also constrain the utilization at 7T of some widely-used 3T ASL techniques, such as pCASL labeling and background suppression. As a consequence, to remain within SAR limits, many UHF ASL implementations employ poor temporal resolutions that are not compatible with the requirements of adequately sampling the hemodynamic stimulus response in fMRI (Ghariq et al., 2012; Luh et al., 2013; Zuo et al., 2013; Bause et al., 2016; Zimmer et al., 2016). For example, the increased power deposition at 7T leads to doubling of the minimum achievable repetition time (TR) of pCASL at 3T, even with low flip-angle readouts (Zuo et al., 2013; Wang et al., 2015). Third, resolution loss along the phase-encoding direction and blurring in echo-planar imaging (EPI) can be observed for readout lengths typically used at lower fields due to the faster T2* decay at UHF (Farzaneh et al., 1990; de Zwart et al., 2006).

Parallel imaging addresses some of these issues by shortening the EPI readout and reducing the effective echo spacing. Thus, using parallel imaging reduces the image blur and geometric distortions due to off-resonance effects (de Zwart et al., 2006), while it facilitates the increase of temporal resolution or brain coverage, reducing the minimum echo time (TE) possible and/or enabling multiple echoes. Acquisition at shorter echo time increases the perfusion-weighted signal

(St Lawrence and Wang, 2005), while multi-echo approaches can better characterize the CBF and BOLD time-courses obtained from the ASL data. Since parallel imaging involves data undersampling, it also leads to decreases in image SNR and, to a smaller extent, temporal SNR (tSNR) (de Zwart et al., 2006).

The improved image SNR at 7T compared to 1.5 and 3T can also be utilized to image the human brain at higher spatial resolution. Pfeuffer and colleagues demonstrated at 7T in humans that functional CBF mapping has superior spatial specificity than the typically-employed gradient-echo (GE) BOLD signal (Pfeuffer et al., 2002). Decreases in the voxel volume lead to proportional reduction in the ASL SNR, which render high-resolution perfusion imaging particularly challenging. In practice, to obtain sufficient SNR in ASL at 3T, low spatial resolutions (above 3 mm isotropic) are utilized. Therefore, only a few studies have shown CBF maps with in-plane resolution below 1.5 mm (Duong et al., 2002; Pfeuffer et al., 2002; Zuo et al., 2013). It is worth noting that these studies utilized diverse acquisition approaches — SE, GE and turbo-FLASH (TFL), but were all performed at 7T, indicating its potential for high-resolution CBF imaging.

In this work, we explore the benefits offered by a custom 7T FAIR QUIPSS II (Wong et al., 1998) variant with optimized labeling RF pulses and dielectric pads (Ivanov et al., 2016a) in comparison to conventional 3T ASL techniques for concurrent (functional) CBF and BOLD imaging. To facilitate a direct across-field comparison, FAIR QUIPSS II was also utilized at 3T. Moreover, other commonly employed ASL techniques, such as PICORE Q2TIPS (Luh et al., 1999) and pCASL, were also acquired at 3T. Furthermore, 2D EPI readouts and partial brain coverage were selected for their widespread availability and temporal efficiency allowing optimal conditions for concurrent functional CBF and BOLD experiments. In a second set of experiments, we additionally investigated the influence of spatial resolution within the aforementioned constraints in a subset of the techniques — pCASL and FAIR at 3T along with the optimized 7T FAIR. The in-plane resolution was varied between 3.0 mm (conventional) and 1.5 mm (high), while dual-echo data was enabled using the GRAPPA (Griswold et al., 2002) parallel imaging approach. In summary, we systematically examined the influence of imaging parameters affecting the ASL signal quality, including field strength, labeling scheme, spatial resolution, echo time, and use of parallel imaging, to determine their utility for achieving high-quality, high-spatial resolution functional CBF and BOLD imaging.

4.3 Materials and Methods

4.3.1 Subjects

Nine healthy volunteers (4 females, age range 24–42) took part in Experiment 1 investigating the effects of ASL sequence choices on perfusion SNR, control SNR and absolute CBF values. Eight healthy volunteers (4 females, age range 24–30) took part in Experiment 2 investigating the effects of spatial resolution and echo time on perfusion SNR, control SNR as well as on CBF and BOLD functional sensitivity. All participants had normal or corrected-to-normal vision. Written informed consent was obtained from all participants according to the approval of the study protocol by the Ethical Committee of the Faculty of Psychology and Neuroscience, Maastricht University.

4.3.2 Data Acquisition

Measurements were performed on a 3T MAGNETOM Prisma Fit and a 7T whole-body research scanner (Siemens Healthcare, Erlangen, Germany) using a 64- and 32-channel head coil (Nova Medical, Wilmington, MA, USA), respectively. In order to improve the labeling efficiency at 7T, two rectangular 18x18 cm^2 dielectric pads with 5 mm thickness (Teeuwisse et al., 2012) were placed on either side of the head at the level of the temporal lobes. At both 3T and 7T, the eye centers were taken as a reference for the magnet isocenter position ($\sim$ level of pons), instead of the typically chosen eyebrows ($\sim$ level of basal ganglia) to minimize B_0 offsets in the labeling region. Furthermore, the brain-feeding arteries were aligned with the B_0 field by adding cushions below the necks of the participants when necessary. These measures were previously shown to improve labeling efficiency at 7T (Ivanov et al., 2016a). Half of the participants were first scanned at the 3T scanner and then immediately transferred to the 7T scanner, and vice versa for the other half of the participants. The two scanners are located within the same building at a walking distance of a few meters.

4.3.3 ASL sequences

All experiments used prototype ASL sequences written in-house. Figure 4.1 schematically shows the analysis pipeline used in this study. In Experiment 1, three ASL schemes, namely FAIR QUIPSS II (Wong et al., 1998), PICORE Q2TIPS (Luh

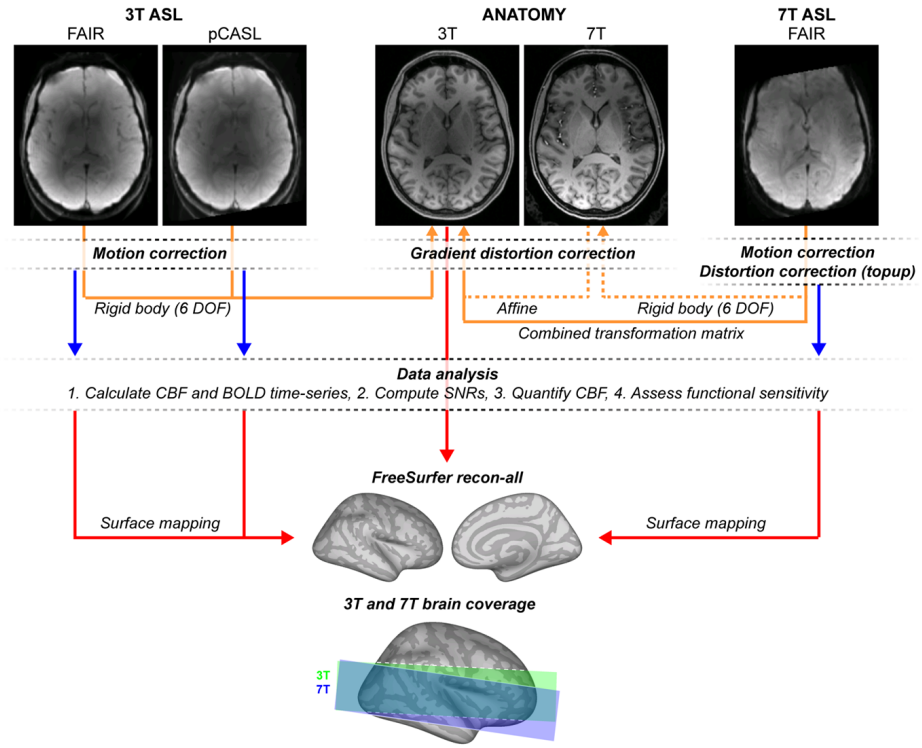

FIGURE 4.1: Schematic illustration of the analysis pipeline.

et al., 1999) and pCASL (Wu et al., 2007; Dai et al., 2008), were acquired at the 3T and one ASL scheme, namely FAIR QUIPSS II, at the 7T. All sequences employed a 2D EPI readout. The labeling at the 7T was achieved using a tr-FOCI inversion pulse (Hurley et al., 2010). For the rest of the paper, we will refer to them with the following abbreviations: 3T FAIR, 3T PQ2T, 3T pCASL and 7T FAIR, respectively. Thirteen slices were acquired with a voxel resolution of 3.0 mm isotropic and angulated in order to cover the visual and auditory cortices. For all schemes, one baseline perfusion run was acquired with 96 time points duration (i.e., 4 min). Further, an additional 3T pCASL run was obtained with identical readout, but labeling duration (τ) and PLD of 1500 and 1580 ms, respectively, in accordance with the ASL white paper recommendations accounting for the distribution of arterial arrival times across the brain (Alsop et al., 2015). The duration of this run (which in the following will be identified with the abbreviation WHITEP) was matched to that of the others, but the TR was 3.6 s, hence resulting in 66

time points. The order of acquisitions at the 3T was pseudo-randomized across the participants. The acquisition parameters were matched as closely as possible across ASL schemes and field strengths (see Table 4.1 for sequence parameters details). At 7T, one M_0 image was acquired using the same sequence but with no magnetization preparation and TR increased to 20 s. At 3T, a separate M_0 acquisition was not necessary since the M_0 acquisition is implemented in the sequence acquisition itself as 1^{st} volume of the run (but only if no parallel imaging is used). At each scanner, one high-resolution anatomical scan was acquired: MPRAGE sequence at 3T with 1.0 mm isotropic resolution (TE/TI/TR = 2.18/1040/2400 ms and 8 degrees flip angle); MP2RAGE sequence at 7T with 0.9 mm isotropic resolution (TE/TI1/TI2/TR = 2.39/900/2750/4500 ms and flip angle 1/2 = 5/3 degrees).

In Experiment 2, two ASL schemes were used at the 3T, namely FAIR QUIPSS II and pCASL, and one ASL scheme at the 7T, namely FAIR QUIPSS II. GRAPPA factor 3 (using the FLEET approach for reference lines acquisition (Polimeni et al., 2016)) was employed for all variants, in order to allow the acquisition of the highest in-plane resolution and the 2^{nd} echo (TEs = 13 and 37 ms) within the fixed TR of 2.5 s. All imaging parameters were matched as closely as possible across ASL schemes and field strengths (see Table 4.2 for details).

For each combination of ASL variant and in-plane resolution in Experiment 2, one functional run of 204 time points (i.e., 8.5 min) was acquired while participants attended to visual stimulation. The stimulation protocol of each run consisted of an initial rest period (gray screen) of 30 s and 8 blocks each consisting of 25 s 8Hz-flickering-checkboard hemifield stimulation followed by 35 s rest. A white cross, always present in the center of the screen, was used as fixation point. In order to stimulate the same portion of the visual field at the 3T and 7T scanner, the size of the projected visual stimulus was adjusted according to the scanner specific stimulus setup to yield 6.5 degrees visual angle stimulation at both scanners. In order to avoid order effects, the acquisition order of the different resolutions was pseudo-randomized across participants as well as that of the ASL schemes at the 3T. For each ASL implementation in Experiment 2, one M_0 image was acquired with a TR increased to 20 s and no magnetization preparation. 7T ASL runs were followed by the acquisition of 3 volumes with the same sequence but with opposite phase-encoding direction to allow for offline distortion correction with the reversed-gradient approach. High-resolution (0.7 mm isotropic) anatomical scans were acquired at each scanner. An

TABLE 4.1: ASL sequence parameters of Experiment 1. All sequences employed a 2D EPI readout. In the central column, all parameter values are reported, while in the left and right column values are displayed only when different among the sequences. The symbol '—' indicates that the corresponding parameter does not apply for that sequence.

	3T FAIR, PQ2T, pCASL	**WHITEP**	**7T FAIR**
In-plane resolution [mm]		**3.0**	
Slice thickness [mm]		3.0	
TR [ms]	2500	3600	2500
TI_1/TI_2 [ms] or	700/1800 (FAIR, PQ2T)	—	700/1800
τ/PLD [ms]	975/1005 (pCASL)	1500/1580	—
Number of slices		13	
TE [ms]		13	
Base resolution		64	
Echo spacing [ms]	0.51	0.51	0.53
Partial Fourier		6/8	
GRAPPA		off	
Slice time acquisition [ms]	30.4	30.4	31.6
Readout duration [ms]	24.48	24.48	25.44

TABLE 4.2: ASL sequence parameters of Experiment 2. In the central column, all parameter values are reported, while in the left and right column values are displayed only when different among the sequences. Values identical across in-plane resolutions are reported only once.

	3T FAIR				**3T pCASL**				**7T FAIR**			
In-plane resolution [mm]	**1.5**	**2.0**	**2.5**	**3.0**	**1.5**	**2.0**	**2.5**	**3.0**	**1.5**	**2.0**	**2.5**	**3.0**
Slice thickness [mm]	3.0											
TR [ms]	2500											
TI_1 or τ [ms]	700				925				700			
TI_2 or PLD [ms]	1800				875				1800			
Number of slices	10				10				12			
TE_1 [ms]	13											
TE_2 [ms]	37											
Delay after 1^{st} echo [ms]	0	6	7	8	0	6	7	8	0	6	7	8
Base resolution	128	96	78	64	128	96	78	64	128	96	78	64
Echo spacing [ms]	0.69	0.68	0.68	0.68	0.69	0.68	0.68	0.68	0.69	0.68	0.68	0.69
Partial Fourier	6/8	6/8	7/8	off	6/8	6/8	7/8	off	6/8	6/8	7/8	off
GRAPPA	3											
Slice time acquisition [ms]	56.8	49.6	47.8	46.0	56.8	49.6	47.8	46.0	54.8	50.5	48.8	47.3
Readout duration [ms]	22.1	16.3	15.6	15.0	22.1	16.3	15.6	15.0	22.1	16.3	15.6	15.2

MPRAGE sequence (TE/TI/TR = 2.11/1040/2400 ms and 8 degree flip angle at 3T and TE/TI/TR = 2.47/1500/3100 ms and 5 degree flip angle) was followed by a proton-density-weighted scan with identical imaging parameters apart from the lack of inversion pulse and shorter TR (1890 ms at 3T and 1660 ms at 7T). This was done to allow correction for receive bias fields in a consistent manner across both scanners (Van de Moortele et al., 2009).

4.3.4 Preprocessing

Motion correction was performed using SPM8 (`http://www.fil.ion.ucl.ac.uk/spm/software/spm8/`). In order to treat all ASL runs equally, each ASL run (and corresponding M_0 image) was realigned independently from the other runs, by realigning all volumes and the M_0 image to the first volume of that run. Also 1^{st} and 2^{nd} echo images of Experiment 2 were treated separately for motion correction as well as for the rest of preprocessing described below. After realignment, 7T ASL runs were distortion corrected using the reversed-gradient approach (Andersson et al., 2003) as implemented in FSL (topup; Smith et al., 2004). The temporal mean of each ASL run was computed and used to calculate the coregistration matrix with the anatomical scan of the corresponding scanning session. Rigid-body coregistration (i.e., using 6 parameters: 3 translation and 3 rotation parameters) between each ASL run and anatomical scan was performed using FLIRT (Jenkinson et al., 2002). In order to align the data from the two different scanners, an affine transformation between 7T and 3T anatomical scan was calculated using RobustRegister (mri_robust_register; Reuter et al., 2010). Prior to that, the anatomical images from both the 3T and 7T sessions were corrected for gradient non-linearities using the vendor-provided software routines. Finally, a mask representing the overlap volume between all ASL slabs (both 3T and 7T sessions, all ASL schemes, and all resolutions for Experiment 2) was computed.

The only exceptions to this pipeline were made for subject 1, 2, and 8 of Experiment 1: For subject 1 and 2, no opposite phase-encoding images were acquired for the 7T ASL run, thus no distortion correction was performed. For subject 8, no 3T anatomy was available due to time constraints during the scanning session. Therefore, ASL data from the 3T session were manually registered to the 7T anatomical scan.

In both Experiment 1 and 2, automatic segmentation of the subject's anatomy was performed on the 3T anatomical scan (except subject 8) using FreeSurfer

(`https://surfer.nmr.mgh.harvard.edu/`) and the recon-all pipeline. The segmentation obtained was visually inspected and no manual corrections were needed for any subject.

4.3.5 Perfusion calculation

Each motion-corrected (and distortion-corrected, in case of 7T) ASL run was separated into control and label time-series. For Experiment 2, only the 1^{st} echo signal was considered. If not further specified, analyses and results concerning Experiment 2 will refer to the 1^{st} echo images for the rest of the paper. Subtraction of the label from the control images yielded a perfusion-weighted time-series and its temporal mean a perfusion-weighted map.

Quantification of the perfusion map was performed according to the model described in (Alsop et al., 2015). The formula for PASL schemes is

$$CBF = \frac{6000 \cdot \lambda \cdot \Delta M \cdot e^{\frac{TI_2}{T1_{blood}}}}{2 \cdot \alpha \cdot TI_1 \cdot M_0} \, [ml/100g/min] \tag{4.1}$$

while for pCASL it is

$$CBF = \frac{6000 \cdot \lambda \cdot \Delta M \cdot e^{\frac{PLD}{T1_{blood}}}}{2 \cdot \alpha \cdot T1_{blood} \cdot M_0 \cdot \left(1 - e^{-\frac{\tau}{T1_{blood}}}\right)} \, [ml/100g/min] \tag{4.2}$$

where ΔM is the (relative) perfusion value (i.e., the temporal mean of the control-label signal intensity difference). TI_2 and PLD values were calculated for each slice, as slices were acquired using a 2D EPI readout with ascending slice order. The constants used in the formulas were: $T1_{blood}$ the longitudinal relaxation time of blood in seconds ($T1_{blood}$ = 1.650 s at 3T (Lu et al., 2004; Zhang et al., 2013); $T1_{blood}$ = 2.100 s at 7T (Dobre et al., 2007; Gardener et al., 2009; Zhang et al., 2013)), λ the brain/blood partition coefficient in ml/g (λ = 0.9 ml/g); and α the labeling efficiency (α = 0.98 for 3T FAIR and 3T PQ2T, α = 0.95 for 7T FAIR, and α = 0.85 for 3T pCASL and WHITEP). Finally, the factor 6000 converts the units from 1/s to ml/100g/min, which is the commonly used physiological unit for CBF.

A mean gray matter (GM) perfusion value was calculated considering all voxels included in i) a CBF-based GM mask, or, alternatively, ii) an anatomically defined GM ribbon mask. The CBF-based GM mask was generated for each ASL acquisition by thresholding the quantitative CBF map between 20 and 120

ml/100g/min, which are physiologically reasonable values for GM CBF (Alsop et al., 2015). Results relative to the anatomically defined GM ribbon mask are nearly identical. Thus, details on methods and results with this latter approach are only reported in the Supplementary Material (section 4.7.2).

Differences in mean GM perfusion values for the different ASL implementations were tested using a one-way repeated measures ANOVA for Experiment 1 and two-way repeated measures ANOVA for Experiment 2. In the first case, ASL scheme (i.e., 3T FAIR, 3T PQ2T, 3T pCASL, WHITEP or 7T FAIR) was the only independent variable, in the latter also voxel volume (e.g. 1.5x1.5x3.0 mm^3, 2.0x2.0x3.0 mm^3, ...) was taken into account as a second independent variable. Repeated measures ANOVA, here and for the rest of the paper, were computed in SPSS using the Greenhouse-Geisser correction (which accounts for unequal variances of the differences between all factor combinations). P-values of pairwise comparisons are reported according to Fisher's Least Significance Differences (LSD, unadjusted probabilities).

Finally, the influence of using parallel imaging (GRAPPA 3) on the baseline perfusion values was tested using the data from Experiment 1 and 2 for the common voxel resolution of 3.0 mm isotropic and the common ASL schemes (3T FAIR, 3T pCASL, and 7T FAIR). A two-way repeated measure ANOVA was computed using ASL scheme as a within-subject factor and parallel imaging as between-subject factor.

4.3.6 SNR measures

The signal quality of the different ASL acquisitions was compared by assessing the perfusion tSNR, control tSNR and perfusion SNR.

Voxel-wise perfusion tSNR was calculated dividing the mean of the perfusion time-series in each voxel by its standard deviation. With the term *control tSNR* we indicate the tSNR of the control time-series, which approximates the tSNR usually measured in standard BOLD fMRI. Control tSNR was computed voxel-wise by dividing the temporal mean of the control time-series by its standard deviation. This quantity was calculated after filtering the control time-series with a nonlinear high-pass filter of 12 TRs using the -bptf option of fslmaths (from the FSL toolbox `http://fsl.fmrib.ox.ac.uk/fsl`).

The perfusion SNR was calculated from the perfusion-weighted map according to the method of (Glover and Lai, 1998; Feinberg et al., 2013), as this definition

of SNR is well suited also for acquisitions using parallel imaging. The perfusion-weighted time-series was divided into two time-series containing either its even- or odd-numbered volumes. A temporal mean was computed for each split time-series, then a sum image and a difference image of the two temporal mean images were calculated. The perfusion SNR is computed as the mean value across the sum image divided by the standard deviation across the difference image. For this calculation, only voxels included in the CBF-based GM mask (described in the previous section 4.3.5 *Perfusion calculation*) were considered. For Experiment 2, all three SNR quantities were calculated also for the 2^{nd} echo time-series.

Differences in perfusion and control tSNR, and perfusion SNR for the different ASL implementations were tested using a one-way repeated measures ANOVA with ASL scheme as factor for Experiment 1 and two-way repeated measures ANOVA with ASL scheme and voxel volume as factors for Experiment 2.

The decreases in the perfusion or control tSNR due to a certain factor, such as use of parallel imaging acceleration, TE or change in voxel volume with respect to a given other reference setting, were obtained using the equation:

$$Loss = \frac{\text{tSNR}_{Ref} - \text{tSNR}_{x}}{\text{tSNR}_{Ref}} \tag{4.3}$$

4.3.7 CBF and BOLD activation

In Experiment 2, the functional sensitivities of CBF and BOLD signal were evaluated by comparing the significant activation detected using the full GLM ASL model (Mumford et al., 2006; Hernandez-Garcia et al., 2010). In this framework, the raw ASL signal is modeled by four predictors:

$$y(t) = \beta_0 + \beta_1 x_1(t) + \beta_2 x_2(t) + \beta_3 x_3(t) + \epsilon(t) \tag{4.4}$$

where $y(t)$ is the signal intensity of a particular voxel at the time point $t = 1, \ldots, n$, the first regressor $x_0(t) = 1$ and its coefficient parameter β_0 represent the baseline MR signal. The second regressor $x_1(t)$ consists of an alternation of $+0.5$ and -0.5 describing the ASL zig-zag signal representing the acquisition of control and label pairs. Thus, its coefficient parameter β_1 represents the relative baseline perfusion. The third regressor $x_2(t)$ and its coefficient β_2 describe the relative CBF signal change due to activation, while the fourth regressor $x_3(t)$ and its coefficient β_3 the BOLD signal change. Finally, $\epsilon(t)$ is the error term. The activation BOLD

predictor x_3 is built convolving a box car function representing the stimulus block with a gamma function in order to account for the hemodynamic response delay. The activation CBF predictor x_2 is obtained by multiplying the activation BOLD predictor x_3 with the baseline CBF predictor x_1.

GLM analysis was performed using FEAT v6.00 (from the FSL toolbox `http://fsl.fmrib.ox.ac.uk/fsl`) including a pre-processing step of high-pass filtering (60.0 s cutoff) and FILM prewhitening. Statistical results were corrected for multiple comparisons using a cluster size approach: voxel Z-statistics was thresholded using a Z threshold of 2.3, then the resulting clusters were thresholded using a cluster p-value threshold of 0.05. Subject 6 was excluded from this analysis as no activation was detected for multiple acquisitions due to excessive motion. For all other subjects, activation maps were masked projecting back into the functional space using the binary mask representing the intersection of all ASL slabs acquired for that subject. In this manner, we excluded that differences in the amount of activation detected were caused by different coverage of the ASL slabs. Thanks to careful positioning of the slices, the intersection across all ASL runs resulted in a consistent portion of the acquired ASL slabs for all subjects (see Figure 4.1 for illustration). Differences in the amount of activated volume detected by the activation CBF and BOLD predictors for the different ASL implementations were tested using a three-way repeated measures ANOVA with signal (CBF or BOLD signal), ASL scheme (3T FAIR, 3T pCASL, or 7T FAIR), and voxel size (corresponding to 1.5, 2.0, 2.5, and 3.0 mm in-plane resolution) as factors. Each subject's activation maps were projected on the cortical surface and averaged for display purposes only.

The full GLM ASL analysis and subsequent statistical evaluation were repeated for the 2^{nd} echo signal of Experiment 2.

4.3.8 Perfusion territories

To assess differences in CBF distribution across the cortex due to the different labeling schemes, we defined perfusion territory masks corresponding to anterior, middle and posterior cerebral arteries (ACA, MCA, and PCA, respectively). The territory definition was based on the atlas provided by (Tatu et al., 2012). To build the territory masks in a semi-automatic way, the FreeSurfer automatic segmentation following the Desikan-Killiany' cortical atlas (Desikan et al., 2006) was used, with the exception of the distinction between occipital and parietal superior

gyrus for which the Destrieux' cortical atlas (Destrieux et al., 2010) was utilized. The segmented GM areas were re-labeled as ACA, MCA or PCA according to the Tatu's atlas. Additionally, left and right distinction was made yielding 6 perfusion territory masks for each subject (see Table 4.5 in Supplementary Material for detailed list of the included areas and Figure 4.6 for an example of the resulting masks). Differences in CBF across perfusion territories were statistically assessed using four-way repeated measures ANOVA with ASL scheme, voxel volume, perfusion territory and laterality as factors for Experiment 2, and three-way repeated measures ANOVA with ASL scheme, perfusion territory and laterality as factors for Experiment 1.

4.4 Results

4.4.1 Perfusion maps

Figure 4.2A shows an axial slice of the quantitative perfusion maps obtained for a representative subject from each ASL acquisition in Experiment 1. Overall, the perfusion maps look similar, however they present some local differences. A clear contrast between GM and WM regions is visible in all maps, but the GM's perfusion signal is generally lower in the 7T FAIR map. The sagittal sinus appears bright for the 3T FAIR scheme, which is a well-known artifact of the labeling geometry causing tagged venous blood to inflow from superior slices (Wong et al., 1997; Cavusoglu et al., 2009). The sparser perfusion map appearance in the posterior part of the brain and the higher values in several isolated voxels in the 3T pCASL are the result of the short PLD used.

Figure 4.2B shows the mean GM CBF averaged across subjects (single-subject and group values are reported in Supplementary Table 4.3). Note that the values indicated by circles in the gray box represent the acquisitions with no parallel imaging utilized. 7T FAIR yields always the lowest GM CBF values, whereas 3T PASL and pCASL yield very similar GM CBF value for the 3.0 mm isotropic measurements without parallel imaging (Experiment 1: '7T FAIR vs 3T FAIR' $p = 0.001$, '7T FAIR vs 3T PQ2T' $p < 0.001$, '7T FAIR vs 3T pCASL' $p = 0.001$, '7T FAIR vs WHITEP' $p = 0.045$), while 3T FAIR delivers the highest GM CBF values for the acquisitions using parallel imaging (GRAPPA 3, Experiment 2: '3T FAIR vs 3T pCASL' $p \leq 0.007$, '3T FAIR vs 7T FAIR' $p \leq 0.002$, '3T pCASL vs 7T FAIR' $p \leq 0.018$ for each voxel resolution). All ASL implementations yield mean GM CBF

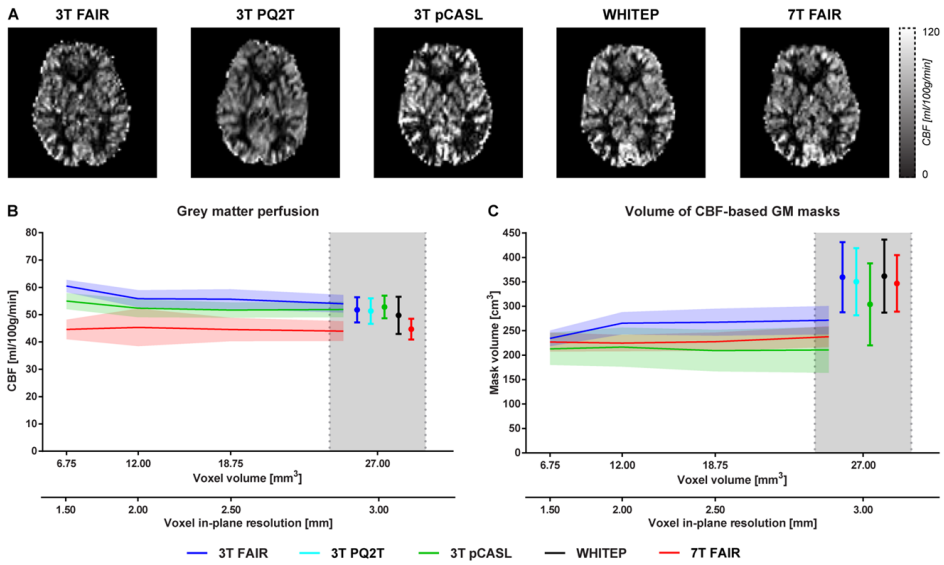

FIGURE 4.2: Quantitative perfusion maps and mean GM CBF values. A) One slice of the perfusion maps obtained with each ASL scheme for a representative subject from Experiment 1; B) Mean GM perfusion values plotted as a function of the voxel volume; C) The volume of the masks (i.e., the number of voxels multiplied by the volume of one voxel) used to compute the mean GM perfusion values plotted as a function of the voxel volume. Shades and error bars represent the standard deviation across the subjects. Two x-axes are shown for clarity in (B) and (C): the first x-axis represents the voxel volume corresponding to each in-plane resolution (indicated in the second x-axis) and a slice thickness of 3.0 mm. Points connected by continuous lines represent the results from Experiment 2, in which the voxel volume varied between 1.5x1.5x3.0 mm^3 and 3.0x3.0x3.0 mm^3 . Results from Experiment 1 with voxels size 3.0x3.0x3.0 mm^3 are indicated by a circle within the gray box and are slightly shifted apart from each other and from the last point of Experiment 2 for visualization purposes.

values in good agreement with the WHITEP implementation. For Experiment 2, the mean GM CBF was generally not influenced by the voxel size: no significant difference was found for 7T FAIR, while a significantly higher mean GM CBF value was found in the 3T FAIR and 3T pCASL case only for the 1.5 mm in-plane resolution ($p \leq 0.001$ for all comparisons with the other voxel resolutions). Figure 4.2C reports the mean volume of the masks used suggesting that smaller volume was detected in the GM mask at 3T for 1.5 mm in-plane resolution due to higher thermal noise, possibly leading to a higher average CBF value, as observed in Figure 4.2B.

Finally, considering the acquisitions with ASL scheme and voxel size common between Experiment 1 and 2 (i.e., the 3T FAIR, 3T pCASL, and the 7T FAIR, all at the 3.0 mm isotropic voxel size), we compared the GM CBF values obtained with or without parallel imaging (i.e., Experiment 2 using GRAPPA 3 vs Experiment 1 without GRAPPA). A significant effect was found for the ASL scheme factor ($p < 0.001$), but not for the parallel imaging factor and the interaction term 'ASL scheme * Parallel imaging' ($p = 0.900$ and $p = 0.179$, respectively). The anatomically defined GM masks yielded nearly identical results (see Supplementary Material 4.7.2).

4.4.2 SNR measures

Temporal SNR

Figure 4.3A shows an example of the perfusion tSNR maps obtained in Experiments 1 and 2. Mean perfusion tSNRs as a function of the voxel size and the use (or not) of parallel imaging are shown in Figure 4.3B (continuous lines and circle symbols). As expected, perfusion tSNR increases with voxel volume and has higher values for acquisitions without parallel imaging. No significant differences in perfusion tSNR were observed for Experiment 1. In contrast, 7T ASL yields higher perfusion tSNR values than 3T at all voxel sizes when parallel imaging is used ($p < 0.001$ for all comparisons; Experiment 2). Within the 3T ASL sequences, pCASL has higher perfusion tSNR than FAIR ($p < 0.001$ for all resolutions). The loss in perfusion tSNR due to the use of GRAPPA factor 3 was largest for 3T FAIR (65%), followed by 58% for 3T pCASL and only 30% for 7T FAIR, indicating the improved parallel imaging performance at 7T. Using the 3.0 mm isotropic resolution from Experiment 2 as a reference, the perfusion tSNR decreased very similarly for 3T FAIR, 3T pCASL and 7T FAIR when increasing the in-plane resolution to 2.5 and 2.0 mm. The differences between 7T and 3T perfusion tSNR loss become large for the 1.5 mm resolution: 46% for 3T vs 33% for 7T. Note that the 7T/3T tSNR ratio, and hence the benefit of 7T, is largest for the highest spatial resolution.

Figure 4.3C shows the tSNR values obtained from the high-pass-filtered control time-series. A consistent trend between different ASL implementations (ASL schemes, voxel volumes, and use or not of parallel imaging) was observed: 7T ASL had higher control tSNR than 3T ASL when parallel imaging was employed

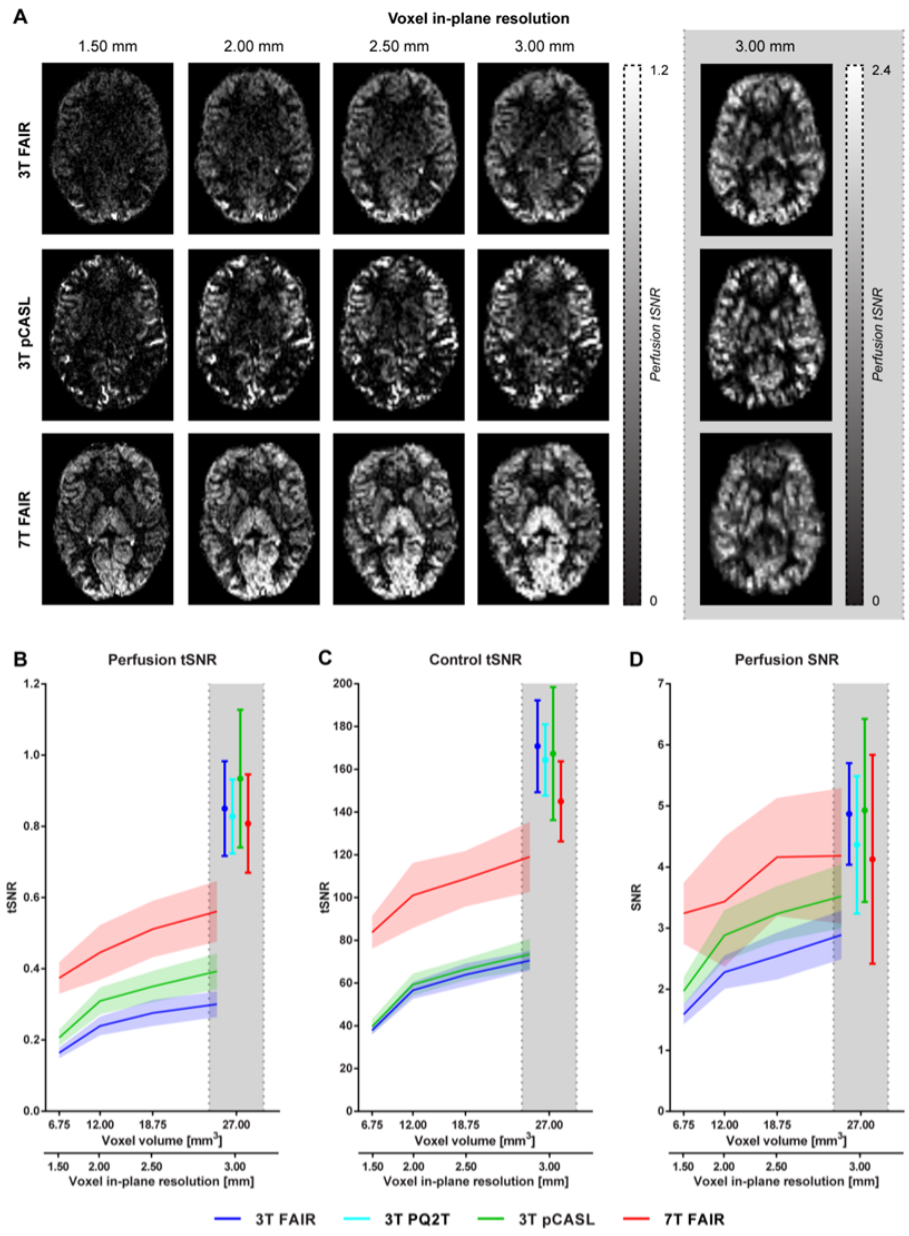

FIGURE 4.3: Perfusion tSNR maps and SNR measures calculated in the CBF-based GM masks. Panel A shows perfusion tSNR maps for two representative subjects from the two experiments. Panel B, C, and D show perfusion tSNR, control tSNR and perfusion SNR, respectively, as functions of voxel volume. Continuous lines and circle symbols indicate measures corresponding to TE = 13 ms for Experiment 2 and 1, respectively. Shades and error bars represent standard deviations across the subjects.

($p < 0.001$ for '7T FAIR vs 3T FAIR' and '7T FAIR vs 3T pCASL'), and lower control tSNR than 3T ASL for implementations without parallel imaging ($p = 0.007$ for '3T FAIR vs 7T FAIR', $p = 0.014$ '3T PQ2T vs 7T FAIR', and $p = 0.031$ for '3T pCASL vs 7T FAIR'). The control tSNR loss due to the GRAPPA factor 3 used was slightly larger for the 3T FAIR than for 3T pCASL (59% vs 56%). In comparison, the control tSNR loss due to use of the same acceleration factor for 7T FAIR was only 18%, further supporting the improved parallel imaging performance at 7T. In Experiment 2, control tSNR increased significantly with increasing voxel volume ($p < 0.001$ for all comparisons). Taking the 3.0 mm isotropic resolution from Experiment 2 as a reference, the control tSNR decreased by a similar amount for all approaches when increasing the in-plane resolution to 2.5 and 2.0 mm. The differences between 7T and 3T control tSNR loss are again largest for the 1.5 mm resolution (30% for 7T FAIR and 46% for both 3T ASL schemes).

Perfusion SNR

Figure 4.3D illustrates the perfusion SNR values. The main findings are similar to those for the perfusion and control tSNR: Perfusion SNR increases with the voxel volume ($p < 0.001$ for the voxel volume factor), and 7T ASL yields higher perfusion SNR values than 3T ASL when parallel imaging is employed ($p = 0.001$ for '7T FAIR vs 3T FAIR' and $p = 0.009$ for '7T FAIR vs 3T pCASL'), whilst slightly lower values when no parallel imaging is used (not significant). Furthermore, 3T pCASL yields higher perfusion SNR than 3T FAIR in case GRAPPA 3 is utilized ($p < 0.001$ for '3T pCASL vs 3T FAIR'). The main difference with respect to the behavior of perfusion and control tSNRs is the reduced difference between the performance with and without parallel imaging: only 3T sequences slightly benefit in terms of perfusion SNR from the absence of in-plane acceleration, while no significant difference is observed for 7T FAIR. Perfusion tSNR, control tSNR and perfusion SNR were computed also using the 2^{nd} echo (i.e., TE = 37 ms) acquired in Experiment 2; results show the expected decrease and are reported in the Supplementary Material (section 4.7.4).

4.4.3 CBF and BOLD activation

The full GLM ASL analysis showed significant activation for both CBF and BOLD signal (see Figure 4.4). The activation clusters were mainly located in the right

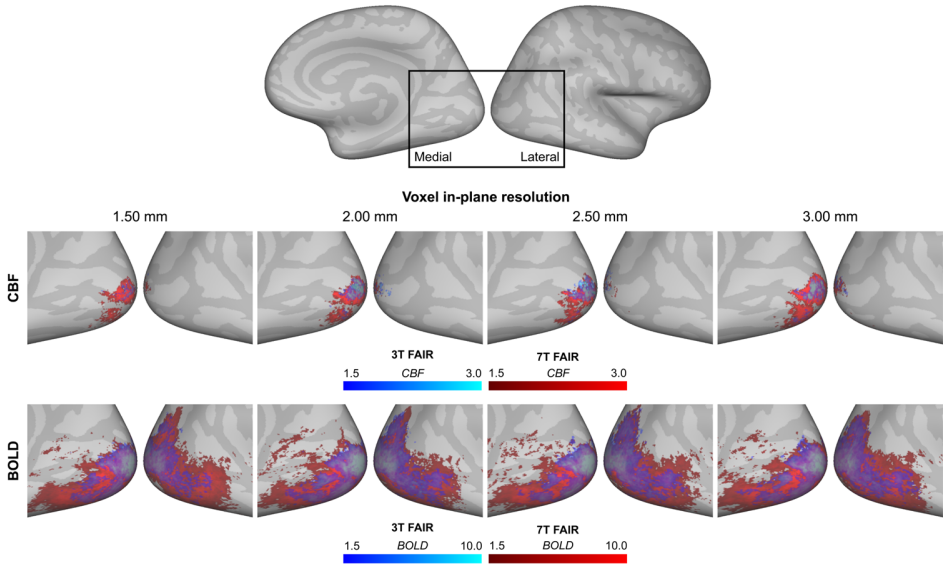

FIGURE 4.4: CBF and BOLD activations. Significant activation detected by the CBF and BOLD predictors of the full GLM ASL model for the different in-plane resolutions: results from 3T FAIR are overlaid on the 7T FAIR ones to facilitate across-field-strength comparison. 3T pCASL activation maps were very similar to those from 3T FAIR and are therefore not shown. The top row shows the CBF activation and the bottom row the BOLD activation. The z-values are color-coded from dark to light blue for 3T FAIR and from dark to light red for 7T FAIR. A different maximum value was chosen for CBF and BOLD (i.e., 3.0 and 10.0, respectively) to better suit their specific dynamic ranges.

early visual cortex (EVC), as expected given the visual stimulation in the left visual field. Note that due to the interpolation when projecting the statistical maps on the cortical surface a lower statistical threshold was chosen for display purposes. The BOLD activation was more widespread and reached higher Z statistics values in line with the higher tSNR observed for the BOLD signal with respect to CBF. In CBF activation, clusters outside EVC were detected only for the lowest resolution investigated (i.e., 3.0 mm isotropic) and 7T FAIR. On the contrary, for BOLD activation, clusters outside EVC were detected for all voxel sizes and more strongly for 7T FAIR. The general increase in the volume activated for increasing voxel sizes and the larger extent of activation detected for 7T FAIR with respect to both 3T acquisitions are evident from the plots in Figure 4.5 (Panels A, and B reffering to CBF and BOLD activation, respectively). The statistical tests resulted in $p < 0.001$ and $p = 0.016$ for the interaction terms 'contrast *

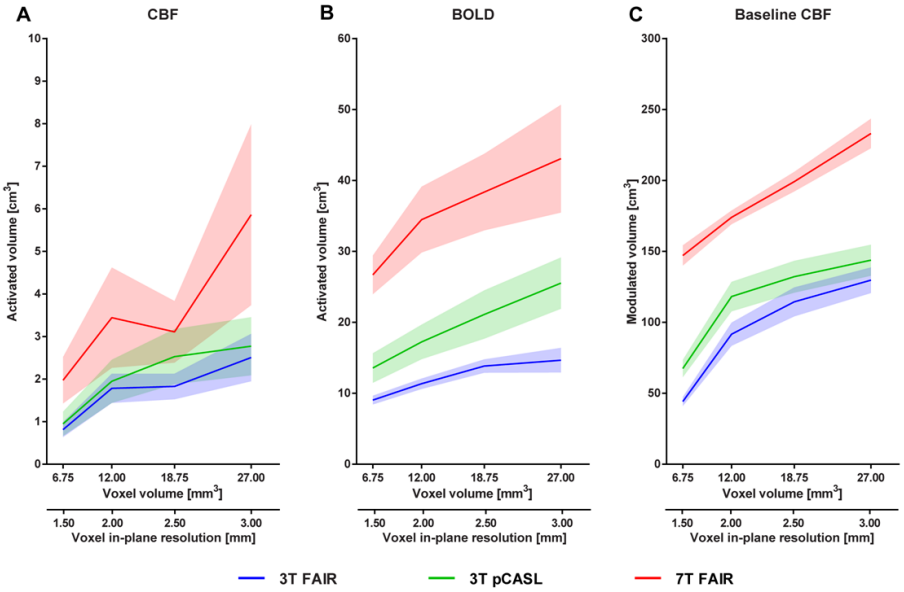

FIGURE 4.5: Results of the full GLM ASL model. Plots A and B show the amount of volume detected as active by the CBF and BOLD predictors, respectively. Plot C shows the amount of GM volume whose time-course was significantly modulated according to the baseline CBF predictor. Lines and shades represent the mean and standard error across the subjects, respectively.

ASL scheme' and 'contrast * voxel volume', respectively. Further, the significant pairwise comparison for the simple two-way interactions were (for CBF/BOLD signal): $p = 0.026/p = 0.055$ for '2.0 mm vs 1.5 mm', $p < 0.001/p = 0.005$ for '2.5 mm vs 1.5 mm', $p = 0.023/p = 0.012$ for '3.0 mm vs 1.5 mm', $p = 0.029$/n.s. for '3.0 mm vs 2.0 mm'; n.s./$p = 0.001$ for '7T FAIR vs 3T FAIR', n.s./$p = 0.001$ for '7T FAIR vs 3T pCASL', n.s./$p = 0.011$ for '3T pCASL vs 3T FAIR', n.s. for all other pairwise comparisons. Finally, 3T PCASL resulted in larger activated volume than 3T FAIR for both CBF and BOLD signal (for CBF/BOLD signal: n.s./$p = 0.011$).

Figure 4.5C shows the volume significantly modulated according to the baseline CBF predictor. The increase in voxel volume resulted in an increase of the modulated volume with a significant effect observed for all comparisons ($p \leq 0.001$ for all comparisons). 7T FAIR surpassed both 3T ASL schemes ($p < 0.001$ for both comparisons), and 3T pCASL was superior to 3T FAIR ($p = 0.002$).

The full GLM analysis was performed also for the 2^{nd} echo data (see Supplementary Material 4.7.5) yielding results in agreement with the BOLD sensitivity depencence on GM T2* at 3T and 7T.

4.4.4 Perfusion territory CBF

Figure 4.6 shows the GM quantitative perfusion values averaged within each territory (which were defined in each individual's anatomical space) and then across the participants to obtain mean and standard error values for all ASL schemes of Experiment 1 and 2. The plots show that the pulsed ASL schemes (i.e., 3T FAIR, 3T PQ2T, and 7T FAIR) result in a higher mean GM CBF in the PCA with respect to ACA and MCA territory, while pCASL leads to a higher mean GM CBF in the MCA territory with respect to the other two, although for the WHITEP implementation the differences among territories are minimized. These different behaviors are reflected in the significance of the interaction 'ASL scheme * perfusion territory' ($p < 0.001$ for both Experiment 1 and 2).

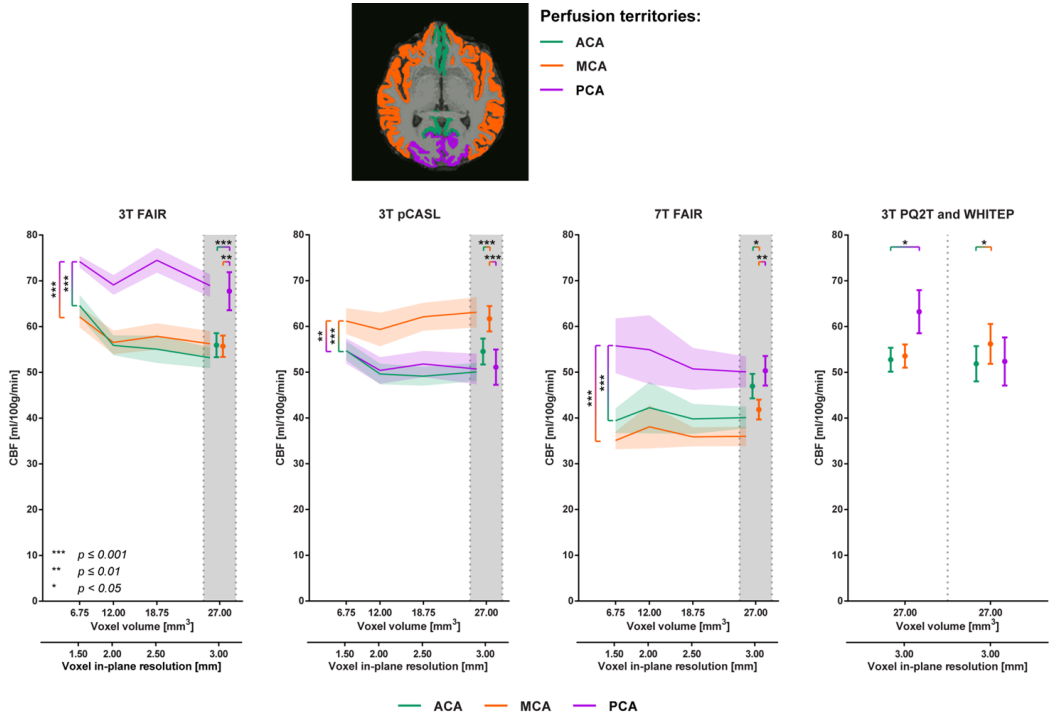

FIGURE 4.6: Perfusion territory GM masks and CBF values. Anatomically-defined perfusion territories are shown (for a representative subject) in the image above the plots: green, orange, and purple masks correspond to ACA, MCA and PCA territories. Territory specific CBF values are shown as a function of voxel volume using the corresponding colors (continuous lines for Experiment 2 and circles for Experiment 1). Note that in this figure, masks and values were shown without differentiating between left and right part of each territory (i.e., contributions from the left and right part of each territory were averaged prior to the display).

4.5 Discussion

In this study, we compared an optimized 7T PASL approach to several 3T ASL techniques with respect to the quality and sensitivity of the concurrently measured CBF and BOLD signals. To this aim, the same labelling scheme, namely FAIR QUIPSS II, was employed at both field strengths and a broad range of measurement parameters was considered: spatial resolution, parallel imaging and echo time. Furthermore, the influence of labeling scheme was also investigated by extending the comparison to alternative 3T ASL techniques, such as PICORE Q2TIPS and pCASL. It is worth reiterating that the focus of this study has been on techniques that allow ASL acquisitions with a short TR (2.5 s) to ensure that a broad range of functional paradigms can be employed.

4.5.1 The influence of magnetic field strength

We observed that 7T ASL has higher perfusion tSNR than 3T ASL for all resolutions and labelling schemes investigated when parallel imaging is used. Control tSNR, TE, bolus length and PLD are the key parameters that determine the perfusion tSNR. When using GRAPPA, the larger control tSNR at 7T with respect to 3T is the main source of improved perfusion tSNR at 7T. In fact, the control tSNR can serve as a metric not only for the effect of field strength, but also for that of parallel imaging and spatial resolution. When parallel imaging is not applied, 7T control tSNR (for 3 mm in-plane resolution) is lower than that of the 3T approaches and the perfusion tSNR is also lower. The higher perfusion tSNR at 7T translates into significantly larger baseline CBF modulated volume and CBF activated volume (Figure 4.5). Figure 4.4 also illustrates the advantages of 7T over 3T for mapping the CBF functional response across the visual areas. These results point out that the image (control) SNR increase (when parallel imaging is used) rather than the longer T1s at 7T plays the decisive role for the advantages of 7T ASL with respect to 3T ASL. The observed 7T benefits for functional BOLD mapping in terms of activation volume and extent are expected and have been reported numerous times (van der Zwaag et al., 2009; Donahue et al., 2011). The smaller portion of the brain significantly activated using CBF than using the BOLD signal is linked to their large difference in tSNR and consequently contrast-to-noise ratio (CNR).

Methods to decrease this disparity in CNR between the two by boosting CBF sensitivity or reducing the BOLD signal sensitivity, like background suppression or alternative readout approaches, are discussed below.

High-quality, high spatial resolution ASL is enabled by parallel imaging through reducing TE and shortening the readout duration to increase the perfusion signal and reduce blurring and susceptibility artifacts. However, the undersampling used in parallel imaging leads to a tSNR decrease, especially in the center of the brain, where the g-factor penalty is largest. This can be observed well in Figure 4.3A, where the 3T perfusion tSNRs using GRAPPA in the subcortical structures are lower across all resolutions in comparison to the non-accelerated 3T and all 7T images. Importantly, it is possible to observe whether this effect is more pronounced for 3T ASL than 7T ASL. The improved parallel imaging performance at 7T has been demonstrated previously (Wiesinger et al., 2004; Keil and Wald, 2013) and is exemplified here by both the perfusion and control tSNR.

The CBF and BOLD signal sensitivity across field strengths also depend on the receive coil array used. In this study, coils with different number of elements were used at the two field strengths because of availability constraints. However, due to the use of 2D EPI, not all coil elements present in the receive array actually contribute to image formation. Thus, we do not expect significant differences when using coils with matching number of elements. Therefore, the influence of the RF coils is considered here as an integral part of the field strength effect on tSNR and parallel imaging performance.

Differences in absolute CBF values between 3T and 7T

Although perfusion quantification is typically not the main point of interest for functional CBF studies, quantification remains a major advantage of ASL compared to BOLD techniques. The obtained quantitative CBF values were largely independent of spatial resolution, parallel imaging, and labeling scheme, but a significantly reduced mean GM CBF at 7T with respect to 3T was found. This finding is in line with those from previous studies at UHF and can have several reasons (St Lawrence and Wang, 2005; Bause et al., 2016). First, labeling efficiency is a direct scaling factor for CBF and using a wrong (overestimated) value for it will lead to underestimation. Second, the temporal duration of the label for PASL sequences is another scaling factor that was assumed, whose overestimation would lead to lower CBF. Both factors may influence CBF estimation

as a function of field strength at the individual and/or group level, since they mainly depend on the B_1^+- and B_0-homogeneity in the labeling region. The field homogeneity at 7T presumably remained lower than typical 3T values, despite the application of dielectric pads and optimized inversion pulses. Substantial technical developments, such as parallel transmission techniques and dedicated labeling coils, might be necessary to achieve labeling performance similar to current 3T scanners. The third potential reason for the lower CBF at 7T relates to the single-compartment model used for its calculation. St. Lawrence and Wang nicely demonstrated that, due to the T2* differences between the capillary blood and tissue compartment, CBF at 7T obtained with this model will be increasingly underestimated with increasing TE (St Lawrence and Wang, 2005). In particular, for the 13 ms TE used here, the underestimation would amount to 10%, which is less than the difference between the 3T and 7T CBF, pointing to the possibility that the two factors described above also play a role.

4.5.2 The effect of spatial resolution

The voxel size decrease is accompanied with a reduction in image (control) SNR and tSNR. The presence of physiological noise prevents improvements in image SNR to translate into commensurate increased tSNR. Increasing the spatial resolution reduces the physiological noise contribution to the time-series variance (Triantafyllou et al., 2011). Consequently, at high spatial resolution, the thermal noise begins dominating the temporal variance, and the relationship between tSNR and voxel volume becomes linear. For which imaging parameters this happens will depend on the field strength, receive coil, acceleration factor, reconstruction approach and echo time. Figure 4.3 suggests that the slope of the GM (t)SNR dependence on the voxel volume at 3T might change between 1.5–2.0 mm and 2.0–3.0 mm in-plane resolution, potentially indicating a change between the thermal noise- and physiological noise-dominated regime. Another indication for this might be the larger tSNR loss between the aforementioned in-plane resolutions at 3T than at 7T. Furthermore, due to the decrease in perfusion and control tSNR with increasing spatial resolution, the CBF and BOLD activation volume and the baseline CBF modulated volume continuously decreased at both fields (Figure 4.5).

There are two main effects that take place when the spatial resolution is increased — the image SNR and time-series tSNR decrease, while the partial voluming between WM, GM and CSF is reduced. The GM thickness across the brain varies between 2 and 4 mm, i.e. it is comparable to the voxel sizes employed in this study. Accordingly, significantly higher CBF at the 1.5 mm in-plane resolution compared to the other resolutions acquired was observed for 3T FAIR and pCASL, but not for 7T FAIR (Figure 4.2B). This suggests the low perfusion tSNR for high spatial resolutions at 3T may be the main cause of this effect: the lower perfusion tSNR at 3T reduces the GM mask volume, eliminating mostly voxels with low perfusion-weighted signal and thus low CBF values. The same trend of higher CBF for the 1.5 mm in-plane resolution at 3T was observed for the values obtained from an anatomically-defined GM mask (Supplementary Figure 4.7), indicating that the reduction of partial voluming at high spatial resolution may also contribute to the result. In particular, the acquisition of smaller voxels leads to increased mean GM CBF, due to the fact that partial voluming with both CSF and white matter decreases the GM voxel's perfusion estimate. The latter effect has been claimed to be one of the reasons for the quantitative discrepancies in CBF observed between ASL and PET (Donahue et al., 2006).

4.5.3 Dual-echo vs single-echo ASL

The perfusion SNR measures and CBF activated volumes all decrease with increasing TE irrespective of the field strength (see Supplementary Figures 4.8 and 4.9). In contrast, the BOLD signal has highest CNR at an echo time equal to the local T2* of tissue. Thus, the TE of the acquisition has a key influence on the CBF and BOLD sensitivities and should be carefully chosen. In the case of a single-echo approach, a compromise between CBF and BOLD sensitivities typically needs to be made. Dual-echo will be beneficial when the second TE is close to the optimal TE for BOLD imaging, while the first TE is as short as possible. Indeed, 3T BOLD activated volume significantly increased for the second TE (independently of labeling scheme and spatial resolution), while the 7T BOLD activated volume obtained using the second TE was smaller than the one obtained at the first TE across all resolutions (Supplementary Figure 4.9B). This latter result can be explained by the significantly shorter T2* at 7T than at 3T, and single-echo ASL should be the preferred option at 7T, especially in the case of high-resolution.

In general, dual-echo will be a viable option at 7T only if low resolution is sufficient, since then the readouts and both TEs can be kept short through the use of partial Fourier or/and parallel imaging. In contrast, dual-echo is the preferred option at 3T across a range of resolutions because the tradeoffs of a single-echo affect negatively both the CBF and BOLD sensitivities.

4.5.4 The influence of labelling scheme

Most functional ASL experiments to date have used PASL, but the recent widespread availability of pCASL has changed this trend, especially for pharmacological and calibrated fMRI studies where rapid temporal changes are uncommon. PCASL has demonstrated higher within-session reproducibility and perfusion tSNR than PASL (Chen et al., 2011). The latter result is also confirmed by our findings. As mentioned above, pCASL allows longer labeling durations, albeit with lower inversion efficiency than PASL schemes. One drawback of pCASL compared to PASL is the lower temporal resolution due to the need for sufficient PLD before the data acquisition to avoid vascular artifacts. In cases when short TRs are required, as in a functional study, a compromise between labeling duration and PLD needs to be found. The tradeoff chosen here is to keep both approximately equal. Our results suggest that a PLD of about 1 second is adequate for the brain regions within the MCA perfusion territory, and potentially for the ones supplied by the ACA, but clearly too short for those perfused by the PCA. Figures 4.2A and 4.3A show perfusion signal inhomogeneities with 'hot spots' of high signal and neighboring dark voxels within the GM belonging to the posterior perfusion territory. This is in contrast to the smooth GM delineation of the PASL and WHITEP images in the same areas. Figure 4.6 depicts the perfusion territories' CBF differences across the approaches tested. PCASL shows significantly higher CBF in the MCA than both the ACA and PCA perfusion territories. In the WHITEP case, relaxation may cause the decreased CBF in the MCA territory, whereas the completed label delivery would increase that of PCA, and only the significant difference between MCA and ACA remains. These results underscore that a PLD of 1.5 s or longer is required to ensure artifact-free perfusion images for pCASL across the entire brain, albeit increasing the current TR for pCASL by 500 ms. Despite the SNR benefits of pCASL at 3T, its practical implementation at 7T is hindered by its high power deposition and sensitivity to field inhomogeneities.

Figure 4.6 indicates that the PASL schemes result in significantly higher CBF in the PCA perfusion territory. The physiological origin of this effect is confirmed by the fact that all PASL schemes employed here exhibit it. One possible explanation of this observation is the higher vascularization of the primary visual cortex. Concurrently, it may also be attributed to the different flow velocities and therefore transit times of the brain-feeding arteries (Guo et al., 2016). The velocity in the vertebral arteries that supply the posterior areas is typically lower than that in the carotids, which deliver blood to the MCA and ACA territories, causing the PCA bolus temporal width to be longer and be delivered later. The QUIPSS and Q2TIPS saturation pulses control the label duration across all feeding vessels, but the label in the PCA perfusion territory has experienced least T1-relaxation and results in stronger perfusion-weighted signal and higher CBF. A wedge-shaped slab-selective inversion pulse, recently proposed by (Guo et al., 2016) may compensate for this effect. It is worth noting that the CBF differences between perfusion territories among labeling schemes reported here should be considered when comparing measurements across ASL approaches, as they constitute a potential confound for baseline and stimulus-induced quantitative CBF values.

The significantly higher perfusion tSNR of pCASL leads to larger CBF activation volume than 3T FAIR, albeit the difference was not statistically significant. It is possible that the higher mean perfusion tSNR for pCASL than PASL, despite the relatively short labelling time, is also related to the insufficient PLD. In comparison, the lack of significant difference in control tSNR between 3T FAIR and pCASL makes the significantly larger BOLD activation volume of the latter appear puzzling. We hypothesize that the disparity in BOLD sensitivity is caused by the difference in longitudinal magnetization between the two, due to the application of presaturation and inversion pulses in FAIR — an effect similar to the application of background suppression.

4.5.5 The effect of background suppression and readout scheme

The omission of background suppression and the use of 2D GE EPI for all the approaches investigated here were dictated by the requirements for short TR, which is incompatible with the utilization of spin-echo approaches due to the SAR restrictions at 7T. Background suppression substantially decreases the tissue signal and physiological noise associated with it and therefore improves the perfusion

SNR. However, since the BOLD sensitivity depends also on the background tissue signal, it is reduced as well. Background suppression is ideally combined with a 3D readout but is also useful in 2D EPI. The key disadvantage, and the main reason for its omission in this study, is its high power deposition making it incompatible with the stringent 7T SAR limits.

The 2D EPI readout was chosen for this functional ASL comparison (and in many other studies), mainly because of its widespread availability, as well as its temporal efficiency and robustness to motion compared to other alternatives, such as segmented 3D SE readouts. Single-shot 3D SE PASL approaches can achieve temporal resolutions similar to the ones used here and may result in improved perfusion measurements (Vidorreta et al., 2013) albeit with reduced BOLD sensitivity. This is a consequence of the SE contrast that removes the static dephasing contribution and hence reduces T2* effects — a strong contributor to the BOLD effect, but harmful for CBF imaging. Partial brain coverage will be required to ensure that the readout does not exceed 300 ms to avoid significant through-plane blurring. The main disadvantage of SE readouts is their high power deposition, which makes them currently impractical at 7T. A readout that shares the robustness to off-resonance effects of the SE approaches, but is well-suited for UHF, is TFL due to the low-flip-angle pulses used (Zuo et al., 2013). TFL has the additional advantage that it is distortion-free, but it is only sensitive to CBF and not the BOLD signal, which makes it unsuitable for their concurrent measurement. 3D EPI readouts (Poser et al., 2010) are well-suited for simultaneous CBF and BOLD measurements (Gai et al., 2011) and have been successfully implemented at 7T (Hall et al., 2010; Ivanov et al., 2016b) because of their low power deposition. 3D EPI outperforms 2D EPI in the thermal noise-dominated regime of high spatial resolutions, but they are comparable in the physiological noise-dominated regime (Poser et al., 2010; Huber et al., 2017). Furthermore, 3D EPI is optimally suited for combination with background suppression, and its main drawback is its current limited availability across sites and platforms. In summary, even though 2D EPI is not the ideal readout for all perfusion applications, it is well-suited for high temporal resolution, simultaneous functional CBF and BOLD mapping, especially in the low spatial resolution regime. The simultaneous multi-slice 2D EPI significantly increases the brain coverage with only a slight SNR penalty (Ivanov et al., 2016a). Sensitivity improvements with comparable temporal resolution, particularly at UHF, may be obtained by the utilization

of 3D EPI with background suppression and/or 3D EPI combined with the double acquisition background suppressed (DABS) FAIR (Wesolowski et al., 2009). At 3T, the combination of a background-suppressed single-shot 3D GRASE readout for the CBF and 2D EPI for BOLD imaging has shown great promise (Fernandez-Seara et al., 2016), but with less than half the temporal resolution achieved here.

4.6 Conclusions

The benefits and challenges of performing ASL studies at 7T compared to 3T were explored. To adequately compare the signal quality and CBF and BOLD signal functional sensitivity, the field strength effect was thoroughly investigated by taking into account also the influence of other acquisition factors, such as parallel imaging, spatial resolution, echo time and labeling scheme. It is demonstrated that high spatial and temporal resolution functional CBF and BOLD imaging can be achieved using the presented single-echo 7T PASL scheme. Its superiority over 3T approaches stems from the higher BOLD sensitivity and improved parallel imaging performance at UHF. At 3T, dual-echo pCASL without acceleration delivers optimal perfusion sensitivity for low spatial and temporal resolution applications.

Bibliography

D. C. Alsop, J. A. Detre, X. Golay, M. Gunther, J. Hendrikse, L. Hernandez-Garcia, H. Lu, B. J. MacIntosh, L. M. Parkes, M. Smits, M. J. van Osch, D. J. Wang, E. C. Wong, and G. Zaharchuk. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. *Magn Reson Med*, 73(1):102–16, 2015.

J. L. Andersson, S. Skare, and J. Ashburner. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage*, 20(2):870–88, 2003.

J. Bause, P. Ehses, C. Mirkes, G. Shajan, K. Scheffler, and R. Pohmann. Quantitative and functional pulsed arterial spin labeling in the human brain at 9.4 T. *Magn Reson Med*, 75(3):1054–63, 2016.

D. P. Bulte, M. Kelly, M. Germuska, J. Xie, M. A. Chappell, T. W. Okell, M. G. Bright, and P. Jezzard. Quantitative measurement of cerebral physiology using respiratory-calibrated MRI. *Neuroimage*, 60(1):582–91, 2012.

R. B. Buxton, V. E. Griffeth, A. B. Simon, F. Moradi, and A. Shmuel. Variability of the coupling of blood flow and oxygen metabolism responses in the brain: a problem for interpreting BOLD studies but potentially a new window on the underlying neural activity. *Front Neurosci*, 8:139, 2014.

M. Cavusoglu, J. Pfeuffer, K. Ugurbil, and K. Uludag. Comparison of pulsed arterial spin labeling encoding schemes and absolute perfusion quantification. *Magn Reson Imaging*, 27(8):1039–45, 2009.

M. Cavusoglu, A. Bartels, B. Yesilyurt, and K. Uludag. Retinotopic maps and hemodynamic delays in the human visual cortex measured using arterial spin labeling. *Neuroimage*, 59(4):4044–54, 2012.

J. J. Chen, H. D. Rosas, and D. H. Salat. Age-associated reductions in cerebral blood flow are independent from regional atrophy. *Neuroimage*, 55(2):468–78, 2011.

W. Dai, D. Garcia, C. de Bazelaire, and D. C. Alsop. Continuous flow-driven inversion for arterial spin labeling using pulsed radio frequency and gradient fields. *Magn Reson Med*, 60(6):1488–97, 2008.

J. A. de Zwart, P. van Gelderen, X. Golay, V. N. Ikonomidou, and J. H. Duyn. Accelerated parallel imaging for functional imaging of the human brain. *NMR Biomed*, 19(3):342–51, 2006.

R. S. Desikan, F. Segonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, 31(3):968–80, 2006.

C. Destrieux, B. Fischl, A. Dale, and E. Halgren. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, 53 (1):1–15, 2010.

J. A. Detre and J. Wang. Technical aspects and utility of fMRI using BOLD and ASL. *Clin Neurophysiol*, 113(5):621–34, 2002.

J. A. Detre, J. S. Leigh, D. S. Williams, and A. P. Koretsky. Perfusion imaging. *Magn Reson Med*, 23(1):37–45, 1992.

J. A. Detre, D. C. Alsop, L. R. Vives, L. Maccotta, J. W. Teener, and E. C. Raps. Noninvasive MRI evaluation of cerebral blood flow in cerebrovascular disease. *Neurology*, 50(3):633–41, 1998.

M. C. Dobre, K. Ugurbil, and M. Marjanska. Determination of blood longitudinal relaxation time (T1) at high magnetic field strengths. *Magn Reson Imaging*, 25 (5):733–5, 2007.

M. J. Donahue, H. Lu, C. K. Jones, J. J. Pekar, and P. C. van Zijl. An account of the discrepancy between MRI and PET cerebral blood flow measures. A high-field MRI investigation. *NMR Biomed*, 19(8):1043–54, 2006.

M. J. Donahue, H. Hoogduin, P. C. van Zijl, P. Jezzard, P. R. Luijten, and J. Hendrikse. Blood oxygenation level-dependent (BOLD) total and extravascular signal changes and DeltaR2* in human visual cortex at 1.5, 3.0 and 7.0 T. *NMR Biomed*, 24(1):25–34, 2011.

T. Q. Duong, D. S. Kim, K. Ugurbil, and S. G. Kim. Localized cerebral blood flow response at submillimeter columnar resolution. *Proc Natl Acad Sci U S A*, 98 (19):10904–9, 2001.

T. Q. Duong, E. Yacoub, G. Adriany, X. Hu, K. Ugurbil, J. T. Vaughan, H. Merkle, and S. G. Kim. High-resolution, spin-echo BOLD, and CBF fMRI at 4 and 7 T. *Magn Reson Med*, 48(4):589–93, 2002.

F. Farzaneh, S. J. Riederer, and N. J. Pelc. Analysis of T2 limitations and off-resonance effects on spatial resolution and artifacts in echo-planar imaging. *Magn Reson Med*, 14(1):123–39, 1990.

D. A. Feinberg, A. Beckett, and L. Chen. Arterial spin labeling with simultaneous multi-slice echo planar imaging. *Magn Reson Med*, 70(6):1500–6, 2013.

M. A. Fernandez-Seara, Z. B. Rodgers, E. K. Englund, and F. W. Wehrli. Calibrated bold fMRI with an optimized ASL-BOLD dual-acquisition sequence. *Neuroimage*, 2016.

N. D. Gai, S. L. Talagala, and J. A. Butman. Whole-brain cerebral blood flow mapping using 3D echo planar imaging and pulsed arterial tagging. *J Magn Reson Imaging*, 33(2):287–95, 2011.

A. G. Gardener and P. Jezzard. Investigating white matter perfusion using optimal sampling strategy arterial spin labeling at 7 Tesla. *Magn Reson Med*, 73(6): 2243–8, 2014.

A. G. Gardener, P. A. Gowland, and S. T. Francis. Implementation of quantitative perfusion imaging using pulsed arterial spin labeling at ultra-high field. *Magn Reson Med*, 61(4):874–82, 2009.

E. Ghariq, W. M. Teeuwisse, A. G. Webb, and M. J. van Osch. Feasibility of pseudocontinuous arterial spin labeling at 7 T with whole-brain coverage. *MAGMA*, 25(2):83–93, 2012.

E. Ghariq, M. A. Chappell, S. Schmid, W. M. Teeuwisse, and M. J. van Osch. Effects of background suppression on the sensitivity of dual-echo arterial spin labeling MRI for BOLD and CBF signal changes. *Neuroimage*, 103:316–22, 2014.

G. H. Glover and S. Lai. Self-navigated spiral fMRI: interleaved versus single-shot. *Magn Reson Med*, 39(3):361–8, 1998.

M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med*, 47(6):1202–10, 2002.

J. Guo, R. B. Buxton, and E. C. Wong. Wedge-shaped slice-selective adiabatic inversion pulse for controlling temporal width of bolus in pulsed arterial spin labeling. *Magn Reson Med*, 76(3):838–47, 2016.

E. L. Hall, P. Gowland, and S. T. Francis. 3D-EPI ASL at ultra high field. *In: Joint Annual Meeting ISMRM-ESMRMB, p 517 Stockholm, Sweden: ISMRM*, 2010.

M. Havlicek, A. Roebroeck, K. Friston, A. Gardumi, D. Ivanov, and K. Uludag. Physiologically informed dynamic causal modeling of fMRI data. *NeuroImage*, 122:355 – 372, 2015.

L. Hernandez-Garcia, H. Jahanian, and D. B. Rowe. Quantitative analysis of arterial spin labeling fMRI data using a general linear model. *Magn Reson Imaging*, 28(7):919–27, 2010.

L. Huber, D. Ivanov, D. A. Handwerker, S. Marrett, M. Guidi, K. Uludag, P. Bandettini, and B. A. Poser. Techniques for blood volume fMRI with VASO: From low-resolution mapping towards sub-millimeter layer-dependent applications. *Neuroimage*, (accepted), 2017.

A. C. Hurley, A. Al-Radaideh, L. Bai, U. Aickelin, R. Coxon, P. Glover, and P. A. Gowland. Tailored RF pulse for magnetization inversion at ultrahigh field. *Magn Reson Med*, 63(1):51–8, 2010.

D. Ivanov, B. A. Poser, L. Huber, J. Pfeuffer, and K. Uludag. Optimization of simultaneous multislice EPI for concurrent functional perfusion and BOLD signal measurements at 7T. *Magn Reson Med*, 2016a.

D. Ivanov, B. A. Poser, S. S. Kashyap, A. Gardumi, L. Huber, and K. Uludag. Submillimeter human brain perfusion. *Proceedings of the High-field Meeting of the ISMRM workshop*, 2016b.

M. Jenkinson, P. Bannister, M. Brady, and S. Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2):825–41, 2002.

B. Keil and L. L. Wald. Massively parallel MRI detector arrays. *J Magn Reson*, 229: 75–89, 2013.

S. N. Krieger, C. J. Gauthier, D. Ivanov, L. Huber, E. Roggenhofer, B. Sehm, R. Turner, and G. F. Egan. Regional reproducibility of calibrated BOLD functional MRI: implications for the study of cognition and plasticity. *Neuroimage*, 101:8–20, 2014.

H. Lu, C. Clingman, X. Golay, and P. C. van Zijl. Determining the longitudinal relaxation time (T1) of blood at 3.0 Tesla. *Magn Reson Med*, 52(3):679–82, 2004.

W. M. Luh, E. C. Wong, P. A. Bandettini, and J. S. Hyde. QUIPSS II with thin-slice TI1 periodic saturation: a method for improving accuracy of quantitative perfusion imaging using pulsed arterial spin labeling. *Magn Reson Med*, 41(6): 1246–54, 1999.

W. M. Luh, E. C. Wong, P. A. Bandettini, B. D. Ward, and J. S. Hyde. Comparison of simultaneously measured perfusion and BOLD signal increases during brain activation with T1-based tissue identification. *Magn Reson Med*, 44(1):137–43, 2000.

W. M. Luh, S. L. Talagala, T. Q. Li, and P. A. Bandettini. Pseudo-continuous arterial spin labeling at 7 T for human brain: estimation and correction for off-resonance effects using a Prescan. *Magn Reson Med*, 69(2):402–10, 2013.

J. A. Mumford, L. Hernandez-Garcia, G. R. Lee, and T. E. Nichols. Estimation efficiency and statistical power in arterial spin labeling fMRI. *Neuroimage*, 33 (1):103–14, 2006.

D. G. Norris. High field human imaging. *J Magn Reson Imaging*, 18(5):519–29, 2003.

J. Pfeuffer, G. Adriany, A. Shmuel, E. Yacoub, P. F. Van De Moortele, X. Hu, and K. Ugurbil. Perfusion-based high-resolution functional imaging in the human brain at 7 Tesla. *Magn Reson Med*, 47(5):903–11, 2002.

R. Pohmann, O. Speck, and K. Scheffler. Signal-to-noise ratio and MR tissue parameters in human brain imaging at 3, 7, and 9.4 tesla using current receive coil arrays. *Magn Reson Med*, 75(2):801–9, 2016.

J. R. Polimeni, H. Bhat, T. Witzel, T. Benner, T. Feiweier, S. J. Inati, V. Renvall, K. Heberlein, and L. L. Wald. Reducing sensitivity losses due to respiration and motion in accelerated echo planar imaging by reordering the autocalibration data acquisition. *Magn Reson Med*, 75(2):665–79, 2016.

B. A. Poser, P. J. Koopmans, T. Witzel, L. L. Wald, and M. Barth. Three dimensional echo-planar imaging at 7 Tesla. *Neuroimage*, 51(1):261–6, 2010.

M. Reuter, H. D. Rosas, and B. Fischl. Highly accurate inverse consistent registration: a robust approach. *Neuroimage*, 53(4):1181–96, 2010.

W. D. Rooney, G. Johnson, X. Li, E. R. Cohen, S. G. Kim, K. Ugurbil, and Jr. Springer, C. S. Magnetic field and tissue dependencies of human brain longitudinal 1H2O relaxation in vivo. *Magn Reson Med*, 57(2):308–18, 2007.

S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J. M. Brady, and P. M. Matthews. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, 23 Suppl 1:S208–19, 2004.

K. S. St Lawrence and J. Wang. Effects of the apparent transverse relaxation time on cerebral blood flow measurements obtained by arterial spin labeling. *Magn Reson Med*, 53(2):425–33, 2005.

L. Tatu, T. Moulin, F. Vuillier, and J. Bogousslavsky. Arterial territories of the human brain. *Front Neurol Neurosci*, 30:99–110, 2012.

W. M. Teeuwisse, A. G. Webb, and M. J. P. van Osch. Arterial spin labeling at ultra-high field: All that glitters is not gold. *International Journal of Imaging Systems and Technology*, 20(1):62–70, 2010.

W. M. Teeuwisse, W. M. Brink, and A. G. Webb. Quantitative assessment of the effects of high-permittivity pads in 7 Tesla MRI of the brain. *Magn Reson Med*, 67(5):1285–93, 2012.

T. Tjandra, J. C. Brooks, P. Figueiredo, R. Wise, P. M. Matthews, and I. Tracey. Quantitative assessment of the reproducibility of functional activation measured with BOLD and MR perfusion imaging: implications for clinical trial design. *Neuroimage*, 27(2):393–401, 2005.

C. Triantafyllou, J. R. Polimeni, and L. L. Wald. Physiological noise and signal-to-noise ratio in fMRI with multi-channel array coils. *Neuroimage*, 55(2):597–606, 2011.

K. Uludag, B. Muller-Bierl, and K. Ugurbil. An integrative model for neuronal activity-induced signal changes for gradient and spin echo functional imaging. *Neuroimage*, 48(1):150–65, 2009.

P. F. Van de Moortele, E. J. Auerbach, C. Olman, E. Yacoub, K. Ugurbil, and S. Moeller. T1 weighted brain images at 7 Tesla unbiased for Proton Density, T2* contrast and RF coil receive B1 sensitivity with simultaneous vessel visualization. *Neuroimage*, 46(2):432–46, 2009.

W. van der Zwaag, S. Francis, K. Head, A. Peters, P. Gowland, P. Morris, and R. Bowtell. fMRI at 1.5, 3 and 7 T: characterising BOLD signal changes. *Neuroimage*, 47(4):1425–34, 2009.

M. Vidorreta, Z. Wang, I. Rodriguez, M. A. Pastor, J. A. Detre, and M. A. Fernandez-Seara. Comparison of 2D and 3D single-shot ASL perfusion fMRI sequences. *Neuroimage*, 66:662–71, 2013.

Y. Wang, S. Moeller, X. Li, A. T. Vu, K. Krasileva, K. Ugurbil, E. Yacoub, and D. J. Wang. Simultaneous multi-slice Turbo-FLASH imaging with CAIPIRINHA for whole brain distortion-free pseudo-continuous arterial spin labeling at 3 and 7 T. *Neuroimage*, 113:279–88, 2015.

R. Wesolowski, P. Gowland, and S. T. Francis. Double Acquisition Background Suppressed (dabs) FAIR at 3 T and 7 T: advantages for simultaneous BOLD and CBF acquisition. *Proceedings 17th Scientific Meeting, International Society for Magnetic Resonance in Medicine*, page 1526, 2009.

F. Wiesinger, P. F. Van de Moortele, G. Adriany, N. De Zanche, K. Ugurbil, and K. P. Pruessmann. Parallel imaging performance as a function of field strength - an experimental investigation using electrodynamic scaling. *Magn Reson Med*, 52(5):953–64, 2004.

D. S. Williams, J. A. Detre, J. S. Leigh, and A. P. Koretsky. Magnetic resonance imaging of perfusion using spin inversion of arterial water. *Proc Natl Acad Sci U S A*, 89(1):212–6, 1992.

E. C. Wong, R. B. Buxton, and L. R. Frank. Implementation of quantitative perfusion imaging techniques for functional brain mapping using pulsed arterial spin labeling. *NMR Biomed*, 10(4-5):237–49, 1997.

E. C. Wong, R. B. Buxton, and L. R. Frank. Quantitative imaging of perfusion using a single subtraction (QUIPSS and QUIPSS II). *Magn Reson Med*, 39(5): 702–8, 1998.

P. J. Wright, O. E. Mougin, J. J. Totman, A. M. Peters, M. J. Brookes, R. Coxon, P. E. Morris, M. Clemence, S. T. Francis, R. W. Bowtell, and P. A. Gowland. Water proton T1 measurements in brain tissue at 7, 3, and 1.5 T using IR-EPI, IR-TSE, and MPRAGE: results and optimization. *MAGMA*, 21(1-2):121–30, 2008.

W. C. Wu, M. Fernandez-Seara, J. A. Detre, F. W. Wehrli, and J. Wang. A theoretical and experimental investigation of the tagging efficiency of pseudocontinuous arterial spin labeling. *Magn Reson Med*, 58(5):1020–7, 2007.

F. Q. Ye, J. A. Frank, D. R. Weinberger, and A. C. McLaughlin. Noise reduction in 3D perfusion imaging by attenuating the static signal in arterial spin tagging (ASSIST). *Magn Reson Med*, 44(1):92–100, 2000.

X. Zhang, E. T. Petersen, E. Ghariq, J. B. De Vis, A. G. Webb, W. M. Teeuwisse, J. Hendrikse, and M. J. van Osch. In vivo blood T1 measurements at 1.5 T, 3 T, and 7 T. *Magn Reson Med*, 70(4):1082–6, 2013.

F. Zimmer, K. O'Brien, S. Bollmann, J. Pfeuffer, K. Heberlein, and M. Barth. Pulsed arterial spin labelling at ultra-high field with a B1+ -optimised adiabatic labelling pulse. *MAGMA*, 29(3):463–73, 2016.

Z. Zuo, R. Wang, Y. Zhuo, R. Xue, K. S. St Lawrence, and D. J. Wang. Turbo-FLASH based arterial spin labeled perfusion MRI at 7 T. *PLoS One*, 8(6):e66612, 2013.

4.7 Supplementary Material

4.7.1 Single-subject and group values of the mean GM CBF

Table 4.3 reports single-subject and group values of the mean GM perfusion calculated using the CBF-based GM masks.

TABLE 4.3: Mean GM perfusion values in the individual CBF-based GM masks and mean and standard deviation (std) across subjects. Values in Table A refer to Experiment 1, those in Table B to Experiment 2.

A

Subj	3T FAIR	3T PQ2T	3T pCASL	WHITEP	7T FAIR
1	51.87	54.02	50.82	47.49	47.49
2	47.48	46.03	52.14	47.55	40.19
3	47.82	46.63	49.00	43.01	44.74
4	56.36	56.22	55.63	53.19	47.37
5	54.44	50.97	53.96	54.06	40.36
6	45.28	45.00	48.23	41.08	38.98
7	51.70	53.13	51.84	—	47.56
8	59.98	58.29	62.09	62.48	47.49
9	51.08	51.62	51.61	49.07	48.17
mean	**51.78**	**51.32**	**52.81**	**49.74**	**44.71**
std	4.64	4.65	4.15	6.80	3.79

B

	3T FAIR				3T pCASL				7T FAIR			
Subj	**1.5 mm**	**2.0 mm**	**2.5 mm**	**3.0 mm**	**1.5 mm**	**2.0 mm**	**2.5 mm**	**3.0 mm**	**1.5 mm**	**2.0 mm**	**2.5 mm**	**3.0 mm**
1	57.35	50.27	50.00	48.68	51.29	47.96	48.58	47.84	40.06	38.03	39.67	38.41
2	60.10	54.45	54.86	54.08	55.16	51.11	50.33	50.27	42.97	43.40	41.73	43.71
3	59.63	56.32	55.14	53.62	54.94	52.91	51.49	53.24	46.52	45.20	46.68	48.10
4	60.38	54.98	56.16	55.24	52.90	50.58	50.41	50.39	47.79	47.88	47.32	45.79
5	63.72	58.58	60.42	55.83	57.74	56.19	53.64	54.36	49.99	59.98	49.73	46.13
6	57.82	53.97	51.54	49.98	51.28	48.63	48.44	49.24	45.21	45.46	48.24	46.68
7	62.51	60.07	59.96	58.14	57.79	55.50	55.13	55.88	44.39	44.75	45.15	44.36
8	62.28	58.23	57.18	56.37	58.67	56.11	55.52	54.06	39.79	38.10	38.00	38.68
mean	**60.47**	**55.86**	**55.66**	**53.99**	**54.97**	**52.37**	**51.69**	**51.91**	**44.59**	**45.35**	**44.57**	**43.98**
std	2.26	3.13	3.66	3.21	2.95	3.31	2.78	2.85	3.58	6.88	4.27	3.62

4.7.2 Mean GM perfusion in the anatomically defined GM mask

Mean GM perfusion values were calculated based on i) a CBF-based GM mask, and, alternatively, ii) an anatomically defined GM mask derived from the anatomical segmentation. Results according to the former approach are presented in the main manuscript; those according to the latter are presented in the following paragraphs. The anatomically defined GM ribbon mask was obtained by binarizing the *ribbon.mgz* output of the recon-all pipeline and transforming it into the functional space of each ASL acquisition. Using both masks and comparing their results allowed us to exclude biases due to arbitrary thresholding in the first case or misregistration between anatomy and functional spaces in the second case.

Figure 4.7 shows the mean GM perfusion values averaged across the subjects for Experiment 1 (circles) and 2 (lines; individual and averaged values are reported in Table 4.4) and plotted as a function of the voxel volume. 3T ASL schemes delivered higher quantitative GM perfusion values than 7T FAIR both with and without using parallel imaging (Experiment 1: $p < 0.001$ for all comparison between 3T ASL schemes and 7T FAIR, except for 'WHITEP vs 7T FAIR' having $p = 0.022$; Experiment 2: $p \leq 0.001$ for '3T FAIR vs 7T FAIR' and '3T pCASL vs 7T FAIR').

Considering the acquisitions with ASL scheme and voxel size common between Experiment 1 and 2 (i.e., the 3T FAIR, 3T pCASL, and the 7T FAIR, all at the 3.0 mm isotropic voxel size) we compared the GM CBF values obtained with or without parallel imaging (i.e., Experiment 2 using GRAPPA 3 and Experiment 1, respectively). We found a significant effect for the ASL scheme factor ($p < 0.001$), and no effect for the parallel imaging factor and the interaction term.

TABLE 4.4: Individual and group mean GM perfusion values in the anatomically defined GM masks. Values refer to Experiment 1 (A) and 2 (B).

A

subj	3T FAIR	3T PQ2T	3T pCASL	WHITEP	7T FAIR
1	57.42	60.99	52.53	48.53	48.72
2	53.11	48.94	60.25	53.07	39.49
3	50.67	48.73	51.11	44.63	41.65
4	66.94	63.24	63.14	60.49	50.40
5	62.88	57.00	64.20	63.69	39.23
6	46.38	43.87	48.10	40.75	32.83
7	58.99	59.02	57.73	—	49.55
8	70.31	65.77	72.79	76.00	49.61
9	58.03	56.40	62.68	53.82	48.00
mean	**58.30**	**55.99**	**59.17**	**55.12**	**44.39**
std	7.65	7.35	7.69	11.36	6.26

B

	3T FAIR				3T pCASL				7T FAIR			
subj	1.5 mm	2.0 mm	2.5 mm	3.0 mm	1.5 mm	2.0 mm	2.5 mm	3.0 mm	1.5 mm	2.0 mm	2.5 mm	3.0 mm
1	56.03	46.64	48.80	46.41	46.74	43.88	49.40	46.97	32.26	30.70	32.13	32.82
2	67.38	57.75	61.08	61.25	56.72	51.37	52.70	54.67	39.52	40.76	37.19	38.62
3	65.62	61.09	59.21	58.19	64.74	60.75	66.96	68.24	44.53	40.48	43.87	47.92
4	64.25	58.66	64.15	62.11	57.23	54.79	56.54	57.91	45.08	43.95	43.32	42.98
5	72.95	64.94	70.18	63.24	64.07	65.25	63.14	65.31	49.69	80.51	49.91	42.48
6	59.69	54.93	52.28	51.47	52.31	45.64	50.05	51.44	40.18	40.10	47.84	43.25
7	69.85	69.51	71.16	66.45	67.39	66.94	68.41	70.35	38.57	38.97	40.88	40.49
8	67.45	62.67	61.04	63.02	63.51	61.63	64.00	59.78	31.36	30.30	28.18	32.45
mean	**65.40**	**59.53**	**60.99**	**59.02**	**59.09**	**56.28**	**58.90**	**59.33**	**40.15**	**43.22**	**40.42**	**40.13**
std	5.44	6.89	7.80	6.76	7.10	8.75	7.67	8.24	6.29	15.84	7.51	5.33

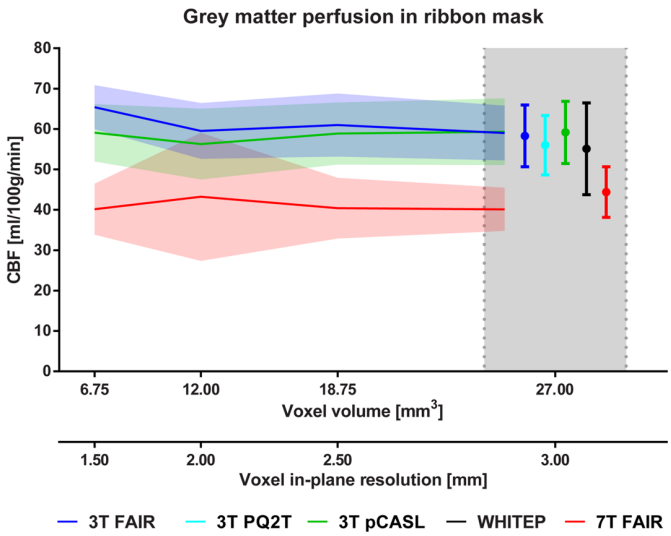

FIGURE 4.7: Mean GM perfusion values in anatomically defined GM masks for Experiment 1 and 2. For each subject, individual GM perfusion values are obtained averaging only voxels in an anatomically defined GM mask. The mean GM perfusion values reported here were obtained averaging across subjects (shades and error bars represent the corresponding standard deviation). Points connected by continuous lines represent the results from Experiment 2 in which the voxel volume was varied between 1.5x1.5x3.0 mm^3 and 3.0x3.0x3.0 mm^3. Results from Experiment 1 when the voxels size was fixed to 3.0x3.0x3.0 mm^3 are indicated by a circle and are slightly shifted apart from each other for visualization purposes. Note that a cyan and black symbol are present only for Experiment 1, as in Experiment 2 no 3T PQ2T and WHITEP acquisitions were performed.

4.7.3 Creation of perfusion territory masks

Table 4.5 reports the list of Desikan-Killiany and Destrieux' areas used to create the six cortical perfusion territories.

4.7.4 SNR measures for 2^{nd} echo data

Perfusion tSNR, control tSNR, and perfusion SNR were computed also using the 2^{nd} echo (i.e., TE = 37 ms) acquired in Experiment 2, and are reported in Figure 4.8 using dashed lines. All three SNR measures importantly decreased for the images acquired at TE = 37 ms with respect to those acquired at TE = 13 ms as expected due to the MR signal decaying with $e^{-TE/T2^*}$. Perfusion SNR and

TABLE 4.5: List of Desikan-Killiany areas used to create the six perfusion territories: left and right ACA, MCA, and PCA. The 'LH' and 'RH' columns report the indices for the areas in the left and right hemisphere while the first column the corresponding name as used by the atlas implemented in Freesurfer (*aparc+aseg.mgz* file). The asterix identifies the additional areas taken from the Destrieux' cortical atlas to distinguish between occipital and parietal superior gyrus.

ACA	**LH**	**RH**
precuneus	1025	2025
isthmuscingulate	1010	2010
posteriorcingulate	1023	2023
paracentral	1017	2017
superiorfrontal	1028	2028
caudal anteriorcingulate	1002	2002
rostral anteriorcingulate	1026	2026
medial orbitofrontal	1014	2014
superiorparietal (intersected with:)	1029	2029
*G_parietal_sup	11127	12127
*S_intrapariet_and_P_trans	11157	12157
MCA	**LH**	**RH**
supramarginal	1031	2031
middletemporal	1015	2015
inferiorparietal	1008	2008
bankssts	1001	2001
superiortemporal	1030	2030
transversetemporal	1034	2034
insula	1035	2035
postcentral	1022	2022
precentral	1024	2024
lateral orbitofrontal	1012	2012
rostral middlefrontal	1027	2027
parsorbitalis	1019	2019
caudal middlefrontal	1003	2003
parstriangularis	1020	2020
parsspercularis	1018	2018
PCA	**LH**	**RH**
lateral occipital	1011	2011
inferiortemporal	1009	2009
fusiform	1007	2007
lingual	1013	2013
pericalcarine	1021	2021
cuneus	1005	2005
superiorparietal (intersected with:)	2029	12120
*G_occipital_sup	11120	12120
*S_oc_sup_and_transverse	11159	12159

tSNR values particularly diminished for 7T FAIR as demonstrated by the absence of significant differences from the corresponding values of 3T pCASL. The loss in perfusion tSNR due to the increase in TE was 64% for 7T FAIR, while it was only 46% for 3T FAIR and 43% for 3T pCASL. Control tSNR values decreased by 58% for 7T FAIR, but remained nevertheless higher than the control tSNR values of 3T pCASL and 3T FAIR (which both declined by 41%). It should be noted that in order to have identical TEs among resolutions and field strength delays up to 8 ms were introduced after the first readout for the larger voxel dimensions (see Table 4.2). Similarly, the first TE could also be shorter than 13 ms by up to 5 ms depending on the in-plane resolution. In result, the TEs for our imaging parameters at 3 mm isotropic resolution without additional delays could have been 8 and 24 ms, respectively — better suited to the T2*s at 7T.

4.7.5 Activation and baseline CBF for 2^{nd} echo data

The full GLM ASL analysis was performed also using the raw ASL signal measured with the 2^{nd} echo (i.e., TE = 37 ms) and the results are presented in Figure 4.9 by the dashed lines. As expected, CBF activation and CBF baseline modulation detection are substantially reduced in comparison to the GLM analysis with the 1^{st} echo (i.e., TE = 13 ms) for all 3T and 7T acquisitions. For the BOLD signal, using the 2^{nd} echo signal yielded larger activation volume for the 3T ASL schemes and comparable activation volume for the 7T ASL acquisition. This behaviour is in line with previous knowledge stating that CBF sensitivity is optimal at short TE values, while BOLD sensitivity is optimal at TE≈T2*. Since the T2* value in the visual cortex at 3T is approximately ∼45 ms, it is very close to the TE_2 used and explains the superior BOLD sensitivity for the 2^{nd} echo. In comparison, the T2* value in the visual cortex is approximately ∼25 ms at 7T, which is halfway between the acquired 1^{st} and 2^{nd} echo. The effect of voxel volume on the detection rate for the 2^{nd} echo followed the trends of the 1^{st} echo with increasing activation volume for increasing voxel volume, although sometimes noisier especially for the CBF activation case.

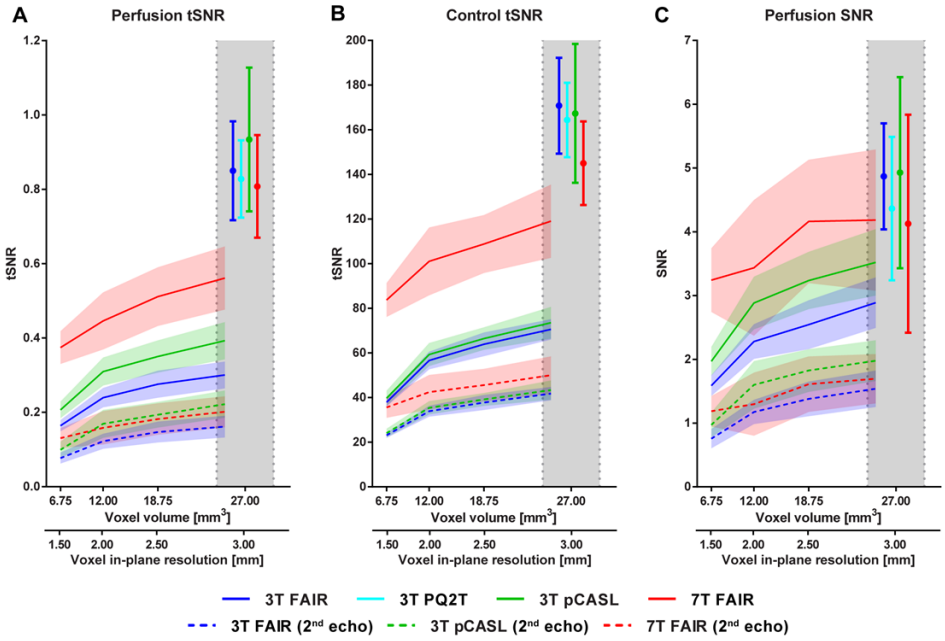

FIGURE 4.8: SNR measures calculated in the CBF-based GM masks. Dashed lines correspond to SNR values calculated for the 2^{nd} echo at TE equal to 37 ms. For comparison purposes, also the SNR values obtained for images acquired at TE equal to 13 ms (i.e., Experiment 1 and 1^{st} echo in Experiment 2), and already discussed in the main manuscript, are reported in the graphs (with circle symbols and continuous lines, respectively). The color coding of the dashed lines is the same as for the continuous lines (blue for 3T FAIR, green for 3T pCASL, and red for 7T FAIR). Shades and error bars represent the standard deviations across the subjects.

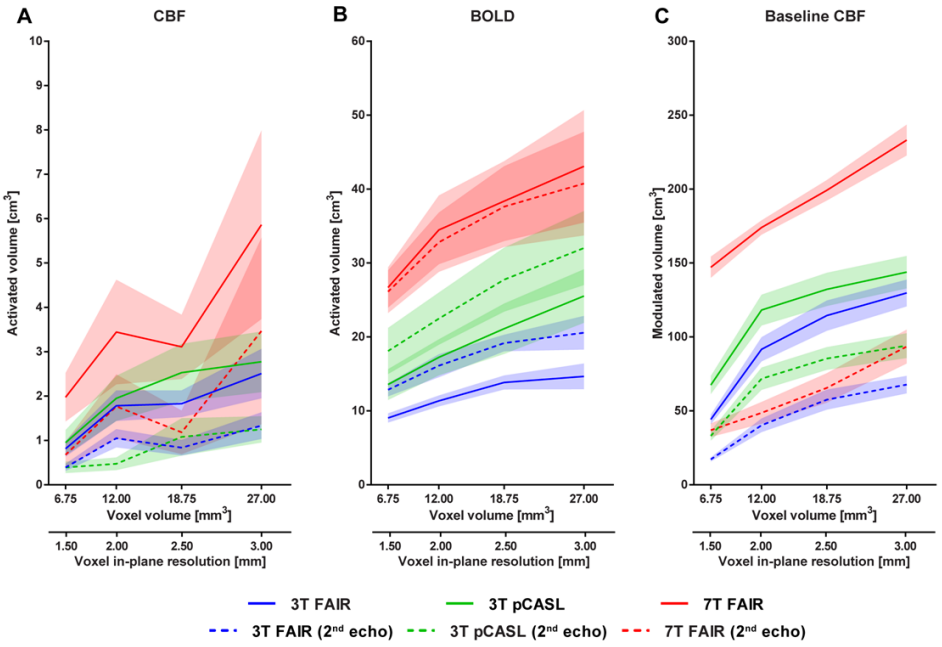

FIGURE 4.9: Results of the full GLM ASL model for the 2^{nd} echo (TE = 37 ms; dashed lines). For comparison purposes also the data corresponding to the TE = 13 ms (i.e., Experiment 1 and 1^{st} echo of Experiment 2) are reported (circles and continuous lines, respectively). Panel A and B show the amount of significantly activated volume as detected by the CBF and BOLD predictors; Panel C shows the amount of volume whose time-course was significantly modulated according to the baseline CBF predictor.

Chapter 5

Summary & General discussion

The research presented in this thesis contributes to the field of fMRI methodology in cognitive neuroscience both in terms of data acquisition and analysis and unveils the mechanisms underlying sound processing in the human auditory cortex.

5.1 Summary

In **Chapter 2**, we addressed an open debate about multivariate pattern analysis (MVPA) in fMRI research. Although, the use of MVPA in fMRI studies has massively increased in the past decade and sometimes even replaced univariate analysis, little is known about the physical and physiological underpinnings of MVPA in fMRI. Different theories have been proposed and investigated, although using a limited range of experiments (visual tasks and 3T fMRI) and yielding conflicting results. Thus, many questions, such as "What is the source of the signal detected by MVPA decoding?", "What is the optimal spatial resolution of acquisition for fMRI data undergoing MVPA?", and "What is the effect of spatial smoothing on MVPA performances?", are still open to debate. In Chapter 2, we have addressed these questions using ultra-high field (7T) fMRI. The primary research aim was to investigate the effect of spatial resolution and smoothing on decoding of speech content (vowels) and speaker identity from auditory cortical responses. The spatial resolution was varied thanks to a novel approach, in which the complex k-space was first reconstructed from the magnitude and phase images acquired at the highest spatial resolution and then downsampled. In addition, data at each resolution were smoothed using a range of 3D Gaussian kernel sizes.

Decoding of vowel and speaker identity was demonstrated to be feasible at 7T for all spatial resolutions and smoothing kernels tested. The effect of spatial smoothing on decoding performances was similar for vowel and speaker decoding: moderate smoothing improved the decoding accuracies, while large smoothing kernels deteriorated them. The effect of spatial resolution, instead, was different across the two decoding tasks: increasing the spatial resolution yielded significantly higher accuracies for vowel decoding, but not for speaker decoding. These results suggest a distinct spatial arrangement of the neural/neuro-vascular sources underlying vowel and speaker representation and processing in the auditory cortex. Moreover, they indicate the need of empirical studies to investigate the optimal acquisition in view of fMRI MVPA as it was shown to differ among

decoding tasks even when using the same stimuli and decoding from the same brain regions.

To characterize the physiological origin of the multivariate information in fMRI, we ranked active voxels according to their likelihood of containing gray matter (micro-vasculature) or large blood vessels (macro-vasculature). We found that discriminability power is distributed over the continuum of tissue type, with a slightly, but significantly, higher contribution from gray matter than from blood vessel voxels. It is expected that this distinction will be further emphasized in going to higher spatial resolution, further reducing partial volume effects. The complex origin of the BOLD signal and its bias in presence of draining veins makes understanding of the role of tissue types in fMRI MVPA even more complicated. Other fMRI techniques, such as cerebral blood flow (CBF) measured with arterial spin labeling (ASL), for instance, have been demonstrated having higher spatial specificity and closer link to the neural activation than GE-EPI BOLD signal. Moreover, CBF is a quantitative physiological measure with straightforward interpretability. It could be therefore beneficial using CBF to study the physiological underpinnings of MVPA and, more generally, the functional organization of the brain. These considerations motivated our **Chapter 3**, in which we used pseudo-continuous ASL at 3T to study the functional organization of the human auditory cortex. In particular, we aimed at mapping tonotopy and voice selective regions in the human auditory cortex using, for the first time, CBF signal instead of the standardly used BOLD signal. Further, we addressed the unresolved issue of the anatomical and functional delineation of the primary auditory core. Finally, we investigated the tissue specificity of CBF and BOLD signal and the possible venous bias of BOLD-based tonotopy.

CBF-based tonotopy showed two main gradients composing a V-shaped pattern of high-low-high preferred frequency centered on Heschl's gyrus and additional frequency gradients in the surrounding regions. The good agreement with BOLD signal-based tonotopy demonstrated the feasibility of CBF-based tonotopy and provided a reciprocal validation of the CBF- and BOLD signal-based findings. Despite the relatively good agreement between CBF- and BOLD-based voice selective maps, two out of five peaks of voice selectivity did not reach significance for CBF. We attributed the missed detection of some voice selective regions and smoother CBF tonotopic maps to the lower SNR of CBF with respect to the BOLD signal. Importantly, quantitative baseline perfusion maps showed a region of high perfusion centered on Heschl's gyrus and corresponding to the main

high-low-high frequency gradients, co-located to the presumed primary auditory core, usually delineated by anatomical landmarks. These observations suggest baseline CBF as a novel anatomical marker for parcellation of the auditory cortex. Finally, we assessed the tissue specificity of CBF and BOLD signal using vein masks computed from SWI images. We found a higher specificity to microvasculature for CBF signal, whilst a venous bias for the BOLD signal. However, our hypothesis that such venous bias of the BOLD signal could be the cause of local mismatches between CBF- and BOLD-based tonotopy did not find significant evidence in the data.

Taking all results of this study together, we showed that both baseline and stimulus-induced CBF is an alternative fMRI approach to the standard BOLD signal to study auditory processing and delineate the functional organization of the human auditory cortex. In future studies, the combination of CBF signal and MVPA analysis could help to further explore the functional auditory cortex and the underpinnings of MVPA fMRI itself.

In Chapter 3, we described the benefits and limitations of ASL fMRI. The latter were primarily identified in the low SNR of CBF signal and motivated the project presented in **Chapter 4**, where different ASL implementations were compared in view of their utilization to simultaneously acquire CBF and BOLD signal in fMRI studies. To that end, the chosen ASL implementations had to fulfill some criteria, such as short TR, TE(s) adequate to both CBF and BOLD acquisition (therefore also no background suppression), (relatively) high spatial resolution, and sufficient brain coverage. We performed one resting-state and one activation study, in which we compared different ASL MRI approaches (FAIR QUIPSS II, PICORE Q2TIPS (only for the resting-state experiment), and pCASL at 3T and FAIR QUIPSS II at 7T). In the resting-state experiment, the voxel size was fixed at 3.0 mm isotropic and no parallel imaging was used. In the activation experiment (consisting of visual stimulation presented in blocks), the acquisitions were repeated at four different resolutions (1.5, 2.0, 2.5, and 3.0 mm in-plane resolution for 3.0 mm slice thickness) and GRAPPA 3 was used. For comparison purposes, one run was acquired with a pCASL implementation tuned accordingly to the recommendations of the ASL white paper (Alsop et al., 2015).

Our results showed that for low spatial resolution (3.0 mm isotropic) and no parallel imaging, 3T FAIR and pCASL yield the highest tSNR and perfusion SNR and are therefore preferable over 3T PQ2T and 7T FAIR. However, increasing spatial resolution and using parallel imaging (GRAPPA) favor the use of 7T FAIR

above all and 3T pCASL above 3T FAIR. Such choice is indicated by tSNR, perfusion SNR and functional sensitivity (assessed as amount of significant activation detected in response to the visual stimuli) measures.

Concerning the quantification of baseline perfusion, 3T ASL schemes resulted in quantitative values in good agreement with those obtained with the "ASL white paper" protocol (chosen as gold standard), providing therefore a validation of the 3T ASL protocols used in our studies. 7T FAIR yielded a significantly lower mean GM perfusion value than 3T ASL schemes. We attributed such discrepancy to differences in the apparent transverse relaxation time of blood and tissue, which are negligible at 3T but not at 7T, and to decreased labelling efficiency at the 7T due to field inhomogeneities.

These results lead to the recommendation to use 3T FAIR or pCASL for studies using low spatial resolution (3.0 mm isotropic and lower) and no parallel imaging, while 7T FAIR for high-resolution and/or studies employing parallel imaging.

5.2 Conclusions and outlook

In conclusion, from a methodological point of view, the studies included in this thesis shed further light on the mechanisms underlying MVPA fMRI, the optimal acquisition and preprocessing of fMRI data undergoing MVPA analysis, the use of stimulus-induced CBF to investigate the functional organization of the brain, the use of baseline CBF to investigate the parcellation of the cortex, and the recommended ASL implementation to employ to achieve these aims. From a neuroscientific point of view, this thesis provides new insights on auditory cortical processing of basic acoustic features (frequency) as well as on the cortical representation of speech content (vowel) and speaker identity. Furthermore, this thesis put forward a novel, simple approach to delineate the primary auditory core.

In the following paragraph, we will comment on specific aspects of the studies, as summarized above, and take the freedom of speculating about possible implications and/or outlooks.

In Chapter 2, we observed different hemodynamic response function (HRF) shapes in different regions of the auditory cortex. For this reason, instead of using a fixed predictor model, we built a range of HRF models with different time-to-peak shifts and chose for each voxel the HRF shift which gave the best fit to

the mean over all trials. This approach represents a first attempt to take into account such HRF variability, however, the flexibility of the models used was quite limited. A more flexible model of HRF shape customizable for each voxel (or brain regions) could result in a more accurate fit of each trial (both training and testing trials) and therefore in an improvement of decoding accuracies.

A range of adaptive models of HRF have been proposed (e.g., Friston et al., 1998; Dale, 1999; Glover, 1999; Woolrich et al., 2004; Lindquist and Wager, 2007), which offer different degrees of flexibility. For instance, in a recent study, Pedregosa et al. (2015) implemented a voxel-by-voxel estimation with a three basis functions model and showed its higher statistical power with respect to other competing models in both encoding and decoding experiments.

Nevertheless, HRF fitting is generally not employed in decoding studies as flexibility of the HRF model comes at the cost of risk of overfitting and resulting in not physiologically plausible hemodynamic shapes, especially in case of low SNR voxels. Therefore, using β-values (or t-values) provided by fitting with "too" flexible models as features of the MVPA algorithm could be, on the contrary, deleterious for the decoding. Moreover, including HRF modelling within MVPA importantly increases its complexity and computational costs as flexible models requires fitting several parameters (and not only one as for GLM with a fixed HRF model). One approach to take into account HRF variability across voxels could be to: 1) estimate the HRF voxel-wise from the training data or an independent experimental paradigm (e.g., using flexible basis functions or a finite response model), 2) describe each voxel's response with a single feature (e.g., its peak amplitude), and then 3) apply the voxel-specific HRFs to derive an estimate of features from the test data (see e.g., Kay et al., 2008a,b; Santoro et al., 2014). Alternatively, one could use multiple features for each voxel including all parameters describing the HRF estimate of that voxel. The rationale behind this approach is that, if the HRF shape itself (and not only its amplitude) carries information about the representation of a certain stimulus, including its parameterization should be informative for the algorithm and beneficial for the decoding. However, combining multiple features in a machine learning algorithm is not trivial and different strategies are possible, among which: the multiple voxel-wise features associated with the different HRF parameters could be concatenated to form a higher dimensional space, or decoding could be performed using multiple kernels corresponding to the multiple features and later recombined.

All different approaches offer advantages, challenges and limitations, thus

theoretical and empirical studies are required in order to develop and evaluate both aforementioned and new approaches to best account for HRF variability in decoding fMRI.

In Chapter 2, we also investigated the tissue specificity of the discriminability power and found that both micro- and macro-vasculature contributed to it. Oversimplifying the issue for a moment, one could think that if the hypothesis of voxel biased sampling is correct, the multivariate information should come from the micro-vasculature (GM), while, if the biased draining region hypothesis is correct, the information should come from the macro-vasculature (draining veins). Our study was conducted with GE-EPI, hence the fMRI signal originated both from micro- and macro-vasculature. As explained in Chapter 3, CBF measured by ASL represents an alternative technique to BOLD fMRI and has the advantage of being localized to the micro-vasculature. Performing a decoding study with ASL could, therefore, allow testing of the two hypotheses and disentangle the origin of the multivariate information. The ability of ASL to simultaneously acquire both CBF and BOLD signal would represent a major strength to this aim, however, under the assumption that both signals had comparable (or at least "enough") SNR. Chapter 3 and 4, unfortunately, speak against the fulfilling of such a requirement by current ASL techniques as the current major drawback of CBF signal is indeed its low SNR (compared to the BOLD signal). In view of future improvements, 7T ASL seems to be the method of choice to this aim because it was found to be the ASL implementation with higher SNR in case of high resolution imaging. High resolution would be indeed another necessary requirement in such project in order to avoid biases due to partial volume effects. Newly introduced 3D ASL acquisitions, such as 3D RARE Stack-Of-Spirals (Ye et al., 2000; Dai et al., 2008; Xu et al., 2010) or 3D GRASE (Gunther et al., 2005), seem also promising in terms of CBF SNR improvement (Vidorreta et al., 2013) and should also be evaluated in further studies.

Baseline perfusion was demonstrated informative on the delineation between primary and secondary areas in auditory cortex (in Chapter 3) and in visual cortex (Weber et al., 2008). Cortical parcellation has been object of study since Brodmann's work (Brodmann, 1909) and has been conducted under the assumption of a close relationship between anatomical and functional properties. Thus, cytoarchitectonics became the leading technique to define structural and functional organization of the brain. Next to it, other techniques such as myeloarchitecture and receptoarchitecture have been used (Smith, 1907; Vogt and Vogt, 1919;

Zilles et al., 2002). With the advent of neuroimaging, cortical parcellation based on myeloarchitecture and (functional) connectivity has gained increasing attention also thanks to their *in vivo* potential (Clark et al., 1992; Barbier et al., 2002; Sigalovsky et al., 2006; Kim et al., 2009; De Martino et al., 2015). Although all mentioned techniques should concur to describe the same functional specialization of cortical areas, the relationship and agreement between the resulting parcellations is still unclear (Cloutman and Lambon Ralph, 2012). Instead of comparing parcellations based on different techniques, recent efforts have been focused on combining them into a multi-modal parcellation approach. In Glasser et al. (2016), a multi-modal approach using both structural (namely, T_1- and T_2-weighted images delivering myelin content and cortical thickness information) and functional (task and resting-state modality) images was combined with machine learning techniques to enable area delineation and thus proving the advantages of integrating information from different modalities. So far, to the best of our knowledge, no multi-modal parcellation method has included perfusion data. However, our findings in Chapter 3 and given the physiological link of CBF with micro-vasculature and metabolic demand of the tissue (Weber et al., 2008), we believe that perfusion signal represents a quite unique marker of both anatomical and functional properties and that it could add relevant complementary information for cortical parcellation.

Multi-modality neuroimaging and machine learning algorithms appear to be two key factors for future developments in the cognitive neuroscience field. A first step in this direction was taken for example by Glasser et al. (2016) who, on the basis of a rich multi-modal database provided by the Human Connectome Project, built an areal classifier able to semi-automatically delineate and identify cortical areas in new individual subjects, even with atypical parcellation. This study also highlights another challenge of the current neuroscientific research: moving from a group-averaging approach to an individual focus and personalized investigation (Dubois and Adolphs, 2016). To this aim, quantitative, robust, and reproducible methods will likely play a major role and open the way for translational applications in clinical and pharmacological research at the single-patient level (Pike, 2012).

Bibliography

D. C. Alsop, J. A. Detre, X. Golay, M. Gunther, J. Hendrikse, L. Hernandez-Garcia, H. Lu, B. J. MacIntosh, L. M. Parkes, M. Smits, M. J. van Osch, D. J. Wang, E. C. Wong, and G. Zaharchuk. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. *Magn Reson Med*, 73(1):spcone, 2015.

E. L. Barbier, S. Marrett, A. Danek, A. Vortmeyer, P. van Gelderen, J. Duyn, P. Bandettini, J. Grafman, and A. P. Koretsky. Imaging cortical anatomy by high-resolution MR at 3.0T: detection of the stripe of Gennari in visual area 17. *Magn Reson Med*, 48(4):735–8, 2002.

K Brodmann. *Localisation in the Cerebral Cortex*. NY: Springer, 1909.

V. P. Clark, E. Courchesne, and M. Grafe. In vivo myeloarchitectonic analysis of human striate and extrastriate cortex using magnetic resonance imaging. *Cereb Cortex*, 2(5):417–24, 1992.

L. L. Cloutman and M. A. Lambon Ralph. Connectivity-based structural and functional parcellation of the human cortex using diffusion imaging and tractography. *Front Neuroanat*, 6:34, 2012.

W. Dai, D. Garcia, C. de Bazelaire, and D. C. Alsop. Continuous flow-driven inversion for arterial spin labeling using pulsed radio frequency and gradient fields. *Magn Reson Med*, 60(6):1488–97, 2008.

A. M. Dale. Optimal experimental design for event-related fMRI. *Hum Brain Mapp*, 8(2-3):109–14, 1999.

F. De Martino, M. Moerel, J. Xu, P. F. van de Moortele, K. Ugurbil, R. Goebel, E. Yacoub, and E. Formisano. High-resolution mapping of myeloarchitecture in vivo: Localization of auditory areas in the human brain. *Cereb Cortex*, 25(10): 3394–405, 2015.

J. Dubois and R. Adolphs. Building a science of individual differences from fMRI. *Trends in Cognitive Sciences*, 20(6):425 – 443, 2016.

K. J. Friston, P. Fletcher, O. Josephs, A. Holmes, M. D. Rugg, and R. Turner. Event-related fMRI: characterizing differential responses. *Neuroimage*, 7(1):30–40, 1998.

M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, S. M. Smith, and D. C. Van Essen. A multi-modal parcellation of human cerebral cortex. *Nature*, 536 (7615):171–8, 2016.

G. H. Glover. Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage*, 9(4):416–29, 1999.

M. Gunther, K. Oshio, and D. A. Feinberg. Single-shot 3D imaging techniques improve arterial spin labeling perfusion measurements. *Magn Reson Med*, 54 (2):491–8, 2005.

K. N. Kay, S. V. David, R. J. Prenger, K. A. Hansen, and J. L. Gallant. Modeling low-frequency fluctuation and hemodynamic response timecourse in event-related fMRI. *Hum Brain Mapp*, 29(2):142–56, 2008a.

K. N. Kay, T. Naselaris, R. J. Prenger, and J. L. Gallant. Identifying natural images from human brain activity. *Nature*, 452(7185):352–5, 2008b.

E. Y. Kim, D. H. Kim, J. H. Chang, E. Yoo, J. W. Lee, and H. J. Park. Triple-layer appearance of Brodmann area 4 at thin-section double inversion-recovery MR imaging. *Radiology*, 250(2):515–22, 2009.

M. A. Lindquist and T. D. Wager. Validity and power in hemodynamic response modeling: a comparison study and a new approach. *Hum Brain Mapp*, 28(8): 764–84, 2007.

F. Pedregosa, M. Eickenberg, P. Ciuciu, B. Thirion, and A. Gramfort. Data-driven HRF estimation for encoding and decoding models. *Neuroimage*, 104:209–20, 2015.

G. Bruce Pike. Quantitative functional MRI: Concepts, issues and future challenges. *NeuroImage*, 62(2):1234 – 1240, 2012.

R. Santoro, M. Moerel, F. De Martino, R. Goebel, K. Ugurbil, E. Yacoub, and E. Formisano. Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Comput Biol*, 10(1):e1003412, 2014.

I. S. Sigalovsky, B. Fischl, and J. R. Melcher. Mapping an intrinsic MR property of gray matter in auditory cortex of living humans: a possible marker for primary cortex and hemispheric differences. *Neuroimage*, 32(4):1524–37, 2006.

G. E. Smith. A new topographical survey of the human cerebral cortex, being an account of the distribution of the anatomically distinct cortical areas and their relationship to the cerebral sulci. *J Anat Physiol*, 41(Pt 4):237–54, 1907.

M. Vidorreta, Z. Wang, I. Rodriguez, M. A. Pastor, J. A. Detre, and M. A. Fernandez-Seara. Comparison of 2D and 3D single-shot ASL perfusion fMRI sequences. *Neuroimage*, 66:662–71, 2013.

C. Vogt and O. Vogt. Allgemeinere Ergebnisse unserer Hirnforschung. *J für Psychologie und Neurologie*, 25:279–461, 1919.

B. Weber, A. L. Keller, J. Reichold, and N. K. Logothetis. The microvascular system of the striate and extrastriate visual cortex of the macaque. *Cereb Cortex*, 18 (10):2318–30, 2008.

M. W. Woolrich, T. E. Behrens, and S. M. Smith. Constrained linear basis sets for HRF modelling using Variational Bayes. *Neuroimage*, 21(4):1748–61, 2004.

G. Xu, H. A. Rowley, G. Wu, D. C. Alsop, A. Shankaranarayanan, M. Dowling, B. T. Christian, T. R. Oakes, and S. C. Johnson. Reliability and precision of pseudo-continuous arterial spin labeling perfusion MRI on 3.0 T and comparison with 15o-water PET in elderly subjects at risk for Alzheimer's disease. *NMR Biomed*, 23(3):286–93, 2010.

F. Q. Ye, J. A. Frank, D. R. Weinberger, and A. C. McLaughlin. Noise reduction in 3D perfusion imaging by attenuating the static signal in arterial spin tagging (ASSIST). *Magn Reson Med*, 44(1):92–100, 2000.

K. Zilles, A. Schleicher, N. Palomero-Gallagher, and K. Amunts. Quantitative analysis of cyto- and receptor architecture of the human brain. In A. W. Toga and J. C. Mazziotta, editors, *Brain Mapping: The Methods (Second Edition)*, pages 573 – 602. Academic Press, San Diego, second edition edition, 2002. ISBN 978-0-12-693019-1.

Knowledge valorization

In this thesis, I presented fundamental research in the field of the cognitive neuroscience. In particular, I used functional magnetic resonance imaging (fMRI) to study auditory processing in the brain (Chapter 2 and 3) and investigated benefits and drawbacks of non-standard fMRI techniques, such as arterial spin labeling (ASL; Chapter 3 and 4) and ultra-high field fMRI (Chapter 2 and 4).

The methodologies and findings presented are, therefore, immediately relevant for researchers using fMRI and/or investigating auditory processing. Moreover, in this chapter, I sketch their possible impact outside fundamental research and how they could be valorized in view of societal benefits. *Knowledge valorization* — also known as *knowledge utilization*, *knowledge transfer*, *science impact*, or *third mission of the universities* — has gained increasing importance in recent years till becoming a central element in science policy (Guston, 2000). In line with many scientists' concern about such new policies (Stemerding & Nahuis, 2014), in my opinion, the relevance of scientific research goes well beyond its immediate or short-term outcomes. Even though its commercial value is not its main drive, past history has demonstrated the complex and unpredictable path from scientific research to societal impact (e.g., penicillin, microwave,...). Thus, scientists have to be aware how important it is for society to valorize the acquired knowledge in a broader sense than the pure seek of knowledge. In the light of these considerations, the following paragraphs present some valorization possibilities of the work illustrated in this thesis.

7T (f)MRI: from research to clinical tool

(F)MRI is a non-invasive imaging technique and therefore represents a highly valuable instrument both for research and clinical purposes. As a consequence, any research project, which challenges the current standards of (f)MRI, has the potential also to significantly contribute future clinical protocols, thus benefiting patient populations. Ultra-high field MRI (i.e., above 4 T) are currently employed almost exclusively in research contexts, while hospitals commonly utilize 1.5 and 3.0 T scanners. In the past few years, more and more evidence on the benefits of ultra-high field (f)MRI has been collected by the scientific community. For example, 7 T MRI was proven superior with respect to 1.5 and 3.0 T in delineating the target areas for deep brain stimulation in Parkinson's disease patients and in depicting the microvasculature of brain tumors (van der Kolk et al, 2013; and references therein). However, strong arguments in favor of maintaining the

current clinical equipment are their lower cost and easier management. Our outcomes of Chapter 2 and 4 contribute to the accumulating evidences that ultra-high field (f)MRI, thanks to the higher signal-to-noise (SNR) ratio and achievable spatial resolution, provide (the scientist here, but potentially also the clinician) richer and therefore more valuable information on cognitive brain processes in healthy subjects and, by generalizing, also in patients. For instance, an MVPA approach similar to the one presented in Chapter 2 (speech content decoding from high-resolution 7T fMRI signal) may be developed to predict which patients with stroke-induced aphasia are likely to recover and which treatment could benefit them the most (inspired by Crinion et al, 2012).

When to use 7T ASL instead of 3T ASL for brain perfusion measurement?

ASL techniques allow measuring brain perfusion in a quantitative and non-invasive manner. Therefore, ASL has a great potential for the diagnostic and following-up of all those pathologies that affect cerebral blood flow (CBF), such as cerebrovascular diseases, dementia and brain tumors. However, medical doctors are quite resistant to introduce ASL measurements as standard diagnostic exam and they still prefer invasive techniques, such as PET, which requires injection of radioactive tracers, or Gd-DTPA contrast agent injection in MRI. Part of this resistance may be attributed to the large variety of ASL methods that have been developed in the past years making it difficult for a practitioner (e.g., clinicians and MRI technicians) to routinely utilize it in applied projects and diagnostics. For this reason, scientists of the ASL community have joined forces to discuss and achieve a consensus on best practices and recommendations for implementation of ASL in standard clinical context, which have been reported in the so-called *ASL white paper* (Alsop et al, 2015). Our study in Chapter 4 moves the discussion about the "best" ASL implementation a step forward comparing 3T ASL implementations not only among each other but also with a 7T ASL variant. This comparison may become extremely relevant once 7T scanners will become a clinically accepted magnetic field strength. Taking into account a wide range of factors influencing ASL performances, we were able to provide concrete recommendations to guide researchers (and clinicians) in setting up study protocols.

ASL, a new non-invasive approach to tackle tinnitus?

In Chapter 3, we used, for the first time, an ASL implementation to study the frequency processing of the auditory cortex in the human brain and we proved its feasibility. Previous studies have demonstrated that CBF signal offers some advantages (such as higher spatial specificity, quantitative measures, and no venous bias) with respect to the standard blood oxygenation level-dependent (BOLD) signal, although with lower SNR. For this reason, one could think to employ ASL to study pathologies, whose origins and/or mechanisms could not be explained by BOLD studies. Given that we validated the functional sensitivity of CBF signal to frequency auditory processing in the human brain, a possible candidate for such exploratory investigation is tinnitus, which is a disease defined by the perception of a sound in the absence of its corresponding external source. Tinnitus is often associated with concentration problems, sleep disturbance, anxiety, depression and extreme fatigue and in its more severe forms may interfere with people's normal life. Tinnitus is affecting about 16–21% of the adult population and 10–15% of the general population seeks medical attention due to its symptoms. Currently, the underlying mechanisms (pathology) of tinnitus are unknown and no cure is available (Maes, 2014; and references therein). It is therefore evident that it is necessary to continue investigating tinnitus, a research context in which ASL fMRI could contribute to new insights and, possibly, to establish a novel clinical treatment.

Acknowledgments

This thesis illustrates the research I conducted in the past 5 years, but it would not be complete without mentioning all the people that made this work possible and, in a way or another, supported me during my PhD time.

First of all, I am very grateful to Kâmil and Elia for believing in me and giving me the great opportunity of doing this PhD. I feel truly honoured that I had the two of you as my supervisors. Thank you Kâmil for introducing me to the world of the cognitive neuroscience, for always finding the time to weekly meet even in your busiest periods, for giving me guidance and at the same time the freedom to explore and learn by myself. Your deep knowledge has been inspiring and your support essential to this PhD. The quality of this work would have not been the same without all our fruitful discussions, your careful reading and thorough feedback. You have been a real mentor to me, thank you! I should thank the fate that brought you, Elia, to that conference in 2011, otherwise I would most likely not be sitting here writing this. Thank you, Elia, for allowing me to join the "Auditory group", I have learned so much from every meeting. Your optimism and encouragement gave me confidence in some crucial moments of my PhD. Thus, thank you not only for your scientific advice, but also for your contagious enthusiasms about science!

Without the collaboration of many people, whose expertise and dedication I greatly esteem, none of my research projects would have been possible. Certainly not without the many hours spent scanning with Dimo. Thank you for setting up and finely tuning all the sequences (for this a big thanks goes also to Ben!). I also appreciated a lot your prompt feedback, your availability to our (sometimes long) discussions whether it was face-to-face or via phone, email, skype, or any possible means of communication. You taught me a lot about MR physics — or, as I would sometimes call it, *MR art* or *MR magic*. Martin, you are my trustworthy HRF and undershoot expert! Thank you for listening and answering to all my BOLD- and CBF-related questions, I have learned so much from you. I truly admire the preciseness you put in each of your projects (and your stunning figures!). When thinking about statistics many doubts arise, but no doubts on whom to ask about…Giancarlo, thank you so much for being always so helpful and your many advices on machine learning and permutation testing. A big thanks to Roy for the numerous tips and tricks about FreeSurfer. Also, your maps and figures reveal you would have quite some talent as a designer, too. Many thanks to Sri for his Linux and PsychoPy expertise and to Federico for sharing his stimulus

code and the fruitful discussions. I would also like to express my gratitude to Christl, Riny, and Eva for all their help with any bureaucratic and logistic matter. Finally, I would like to thank all reading committee and corona members.

A PhD is not only experiments, analyses, and paper writing. It's also office sharing, lunches, coffee breaks, conferences, and even beach volleyball (thanks João for this last one! :)). Above all, I met amazing people and I have now the honor to call some of them friends.

Inge, we shared so many adventures (and pictures!) and I am glad to have you on my side also for this very special one! I admire your cheerful and proactive spirit, you always see the bright side of everything. Thank you for always listening and advising me especially in this last stressful period. I will never forget when you literally rescued me from Charleroi, thank you again so much for that! I loved getting lost (well, I was lost, not you...) in the forest of Sint-Pietersberg with you and Lara (she left us too early, but I like to think she is now running free and happy in the Elysian Fields). You were also the perfect back-packing companion: our trip to Japan was simply amazing!! Where shall we travel next?!?

Gesa, it was great pleasure and honour to be your paranymph just about a year ago and, now that roles are reversed, I am so happy to have you on my side. Your determination and persistency have been of great example to me. Thank you for all your support, for passing by the office, for our lunches together whenever your countless teaching hours allow you, and for always surprising me with so delicious birthday cakes!

Felix, you probably are the most special officemate I will ever encounter in my life. I doubt anybody will ever interrupt my work again to ask: "Do you think that if we go on the dark side of the moon, the universe will appear to our eyes as dark or bright?". It was great fun to wildly speculate about the weirdest existential questions you could come up with in the middle of a working afternoon. Thank you for listening to all my frustrations, never downplaying them and still always making fun of me ("Guess what?" "You were fired!" "No, I was not..."). I admire your acumen and critical thinking, your perspective on things always make me reflect.

After a "Gesa" and a "Felix" there is also a "Gesa & Felix". Guys, you are definitively the best cooks I ever met! No matter which cuisine is on the menu, it is always spot-on. I enjoyed so much all our cooking events accompanied by sometimes light and sometimes serious chats, a lot of laughing and a few tears, and often ending watching a nice movie, an episode of Sherlock, or...the Great

British Menu! You always welcomed me at your place, it felt a bit like family and this means a lot to me. Our road trip from Maastricht to Cologne-Düsseldorf-Nijmegen-Efteling and back, our Italian trip to Pisa and Florence, the theater, the bike trips, and even gardening…we have done so many things together and still I wish we would have done more! More importantly, no matter where we will be in the next years, I hope we will always find the way to meet and keep doing plenty of things together.

The last months of this PhD were not easy, but they would have been a lot harder without sharing our daily stories, challenges, (mis)adventures, frustrations, small and big achievements…Kiki, the pot of gold is waiting for us at the end of the rainbow! I truly enjoyed every single line of our almost daily messages — full of :) ;) lol T.T XD in line with our emotional rollercoaster and some Dutch words of wisdom (above all "De aanhouder wint" has a special place in my memories). Sorry that I infected you with my meteoropathy, but I just felt so much like sharing my irrepressible happiness about sun shining and blue sky. I wish you all the best for the end of your PhD!

I feel very lucky for all the workshops and conferences I had the opportunity to attend during my PhD. If I should name two of them, I would probably pick the Perfusion ISMRM Workshop in Amsterdam, which was my very first workshop and initiated me to the world of perfusion fMRI, and OHBM in Hawaii. The latter of course for the content and not for the location of the conference, but since the trip was so long, it seemed unreasonable not to prolong the visit with a holiday…Aline, Anne, Gorka, Joost, Lars, Michael, Matteo, Martin, Thessely, and Vittoria, it was amazing discovering the gorgeous nature of Kauai and Maui islands together with you.

Talking about traveling, thank you Sanae for all the tips about Japan, I fell in love with such a fascinating country! During all these years I enjoyed a lot our chats and I love how you always keep me updated about your trips or Emma's achievements. I hope you will visit me in Italy and I would love to visit you in Japan one day!

And when I was not the one traveling, Martin and Thes, it was always great pleasure to take care of Toby in your absence. He is such a cheerful and sweet dog, I will miss dog-sitting him once in a while.

Emma, we shared our first experience as paranymphs and it was great fun to shoot Gesa's PhD movie. Thank you (also to Greg!) for inviting me over for my first ever Thanksgiving dinner, it was wonderful!

Judith E., how would I have done without your prompt help figuring out any kind of BrainVoyager tricks?! Thank you! Also for the delicious baked treats and the fun after-work evenings at Thembi.

I know a lot of people like home-office, but I always preferred working at uni. Certainly this was also because of the enjoyable office atmosphere: thank you Felix, Christianne, Matteo, and Evi for that! Matteo, you and your Nespresso bag entering the office door every morning at 11.00 am with a glowing "Buongiorno!" definitively deserve a special mention.

Roberta, too bad our times in Maastricht overlapped only partially, but I enjoyed chatting with you every time our busy schedules allowed us to meet for a coffee in the city center.

I guess you figured out by now how much I enjoyed working at the CN department. For this, I would like to thank also Amaia (Workout Mix on Sunday morning was exhausting but super cool!), Anke, Britta, Faruk, Franzi and Thomas (thank you also for the kind ride on my way back from Berlin!), Gojko, Helen (thank you for all your delicateness and that time you even left me a chocolate muffin on my desk!), Jeannette, Job, Joel, Lukas, Marin, Mario A., Mario S., Martin F., Martha, Michelle, Milene, Minye, Nina, Peter, Rosanne, Tahnée, Tom, Valentin and many other (former) colleague. You made a wonderful work environment!

Maastricht life also meant Hertogsingel housemates: thank you Gojko, Jasmina, Lars, Shanice, and Evi for the nice atmosphere and the convivial brunches, drinks, and dinners.

My period in Maastricht would have not been the same without (quite) some time invested in my main passion: volleyball. I am grateful to Ata Trento Volley for allowing me to play abroad and to Fyrfad (a special thanks to Aniek, Leonie, Sanne, Sara, and Wai-Yan) for welcoming me in the Dutch volleyball — winning the championship and playing the GNSK together was great fun! Thanks to Kivola Riemst (Anne, Aukje, Kim, Sanne, Sara, Sofie, all other teammates and our most loyal supporters; Jef, Bart, and Maddy) I also discovered the Belgian volleyball. Winning the play-offs of the B League and playing in the Belgian A League were extraordinary experiences, whose memories I will always jealously guard.

Un sincero grazie va anche a chi mi ha sempre incoraggiata a distanza. Ari, Daniel, Fede, Guido, Lollo, Marco, Marty, Nene, Valeria e Vally, grazie per rispondere sempre presente a ogni mio rientro a Trento. Che si tratti di un aperitivo in

città, una cena o una passeggiata in montagna, non mancate mai l'occasione per rivederci e aggiornarci e questo mi riempie il cuore.

Merci à mes amis Fra & Lore, Giulia & Mattia. Parigi ci ha fatto conoscere e, anche se ora vorrei ci fossero molti meno chilometri di distanza a separarci, adoro i nostri ritrovi di capodanno (e non solo) in giro per l'Italia e l'Europa. A quando e dove il prossimo ritrovo?!

Un abbraccio e un grazie ad Andrea. "Frate", grazie per i tuoi incoraggiamenti e per farmi ridere delle mie mille preoccupazioni con la tua inconfondibile ironia. Sono troppo felice per te e Sabry e vi auguro tutto il bene del mondo per il vostro futuro insieme! Grazie anche per tutto il calore con cui mi accogliete ogni volta che passo da Trento, non vedo l'ora che ritrovi e cene insieme diventino un appuntamento frequente e non un evento raro.

Infine, il più grande di tutti i grazie va ai miei genitori. Mamma e papà, grazie per accogliermi a braccia aperte (e fare anche il conto alla rovescia!) a ogni mio rientro, grazie per tutti i chilometri che avete fatto per me tra aeroporti e "vacanze" olandesi (penso abbiate perso il conto già da un bel po'...), grazie per il vostro sostegno incondizionato e la vostra più totale fiducia. Anche se la distanza ci separa, il vostro affetto, i vostri insegnamenti e i valori che mi avete trasmesso mi accompagnano sempre. Per questo, ogni mio merito è anche vostro e vi sarò per sempre immensamente grata.

Curriculum Vitae

Anna Gardumi was born in Trento, Italy, on September 23^{rd} 1987. After graduating from high school at Liceo Scientifico Galileo Galilei in Trento in 2006, she studied Physics at the University of Trento. She successfully completed the Honors Program in Physics (PAF certificate) and obtained her Bachelor's Degree cum laude in September 2009. She continued her studies enrolling for the Master program in Biological and Medical Physics. Thanks to an LLP Erasmus scholarship, Anna moved to Paris, France, in August 2010 and attended the Master II of Physique Médicale at the Université Paris-Sud XI. She conducted a ten months internship under the supervision of Dr. Aurélie Desbrée at the Institute de Radioprotectione et de Sûreté Nucléaire in Paris. In December 2011, Anna defended her Master thesis on dosimetry of radiopharmaceuticals in paediatric nuclear medicine and obtained her Master's Degree cum laude in Physics at the University of Trento. In February 2012, she started her PhD program at the Department of Cognitive Neuroscience at Maastricht University, where she conducted her research under the supervision of Dr. Kâmil Uludağ and Prof. Elia Formisano. Currently, Anna is working within the same group as a postdoctoral researcher.

Publications

Peer reviewed journal articles

Gardumi, A., Ivanov, D., Havlicek, M., Formisano, E., and Uludağ, K. (2016). Tonotopic maps in human auditory cortex using arterial spin labeling. *Human Brain Mapping* (In press)

Gardumi, A., Ivanov, D., Hausfeld, L., Valente, G., Formisano, E., and Uludağ, K. (2016). The effect of spatial resolution on decoding accuracy in fMRI multivariate pattern analysis. *NeuroImage 132*, 32-42

Havlicek, M., Roebroeck, A., Friston, K., **Gardumi, A.**, Ivanov., D., and Uludağ, K. (2015). Physiologically informed dynamic causal modeling of fMRI data. *NeuroImage 122*, 355-372

Broggio, D., Moignier, A., Ben Brahim, K., **Gardumi, A.**, Grandgirard, N., Pierrat, N., Chea, M., Derreumaux, S., Desbrée, A., Boisserie, G., Aubert, B., Mazeron, J-J., and Franck, D. (2013). Comparison of organs' shapes with geometric and Zernike 3D moments. *Computer methods and programs in biomedicine 111* (3), 740-754

Hadid, L., **Gardumi, A.**, and Desbrée, A. (2013). Evaluation of absorbed and effective doses to patients from radiopharmaceuticals using ICRP 110 reference computational phantoms and ICRP 103 formulation. *Radiation protection dosimetry 156* (2), 141-159

Gardumi, A., Farah, J., and Desbrée, A. (2013). Creation of ORNL NURBS-based phantoms: evaluation of the volume effect on absorbed doses from radiopharmaceuticals. *Radiation protection dosimetry 153* (3), 273-281

Havlicek, M., Roebroeck, A., Friston, K., **Gardumi, A.**, Ivanov., D., and Uludağ, K. (Under review). On the importance of modeling fMRI transients when estimating effective connectivity: a dynamic causal modeling study using ASL data.

Ivanov*, D., **Gardumi*, A.**, Haast, R., Pfeuffer, J., Poser, B.A., and Uludağ, K. (Under review). Comparison of 3 and 7 Tesla ASL techniques for concurrent functional perfusion and BOLD studies.

*Authors contributed equally to the work

Conference contribution

Gardumi, A., Ivanov, D., Haast, R., Pfeuffer, J., Poser, B.A., and Uludağ, K. (Accepted). The influence of acquisition parameters on CBF and BOLD sensitivity in 3 and 7 Tesla ASL fMRI. *25^{th} Annual meeting of ISMRM*

Ivanov, D., Pfeuffer, J., **Gardumi, A.**, Uludağ, K., and Poser, B.A. (Accepted). 2D CAIPIRINHA improves accelerated 3D GRASE ASL. *25^{th} Annual meeting of ISMRM*

Ivanov, D., Kashyap, S., **Gardumi, A.**, Poser, B.A., and Uludağ, K. (2016) Sub-millimeter human brain perfusion maps using arterial spin labeling at 3 and 7 Tesla. *High-field workshop ISMRM*

Ivanov, D., **Gardumi, A.**, Poser, B.A., Pfeuffer, J., and Uludağ, K. (2015). Comparison of 3 and 7 Tesla arterial spin labeling techniques for simultaneous functional perfusion and BOLD MRI studies. *23^{rd} Annual meeting of ISMRM*

Gardumi, A., Ivanov, D., Valente, G., De Martino, F., Formisano, E., and Uludağ, K. (2015). Tonotopy in the human auditory cortex using Arterial Spin Labeling. *Annual meeting of OHBM*

Havlicek, M., Roebroeck, A., Friston, K., Ivanov., D., **Gardumi, A.**, and Uludağ, K. (2014). New physiological framework for dynamic causal modeling of fMRI data. *Annual meeting of OHBM*

Havlicek, M., Roebroeck, A., Friston, K., Ivanov., D., **Gardumi, A.**, and Uludağ, K. (2013). Dynamic Causal Modeling for Arterial Spin Labeling data. *Annual meeting of OHBM*