universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Evaluation of a minimal assumption approach for human brainstem fMRI"

verfasst von / submitted by

Alexandra Weber BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 834
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Molekulare Biologie
Betreut von / Supervisor:	Beisteiner, Roland, Ao.Univ.-Prof. Dr.med.univ.

Acknowledgements

I would first like to thank my supervisor Roland Beisteiner from the Medical University of Vienna, Department of Neurology, study group clinical fMRI, for giving me the opportunity to write my master thesis at his laboratory. He allowed me to independently write this paper, but supported me with ideas and tips whenever I needed it.

I would also like to thank my colleagues of the study group: Eva Matt, Florian Ph.S Fischmeister, Tuna Aslan, Ahmad Amini and Tabea Philippi-Novak, who made working in this group a precious and fun experience. Especially, I am thankful for all the support Eva Matt gave me during this time! She was my mentor and always answered all of my questions with incredible patience and a smile on her face. She helped me to understand the data and statistical analyses and continuously supported me with very clever ideas. I also want to thank Florian Ph.S Fischmeister, who contributed smart ideas and was always ready to listen when I was struggling with data.

Further I want to thank all the persons who were willing to be subjects for this project. Without them I couldn't have written this thesis.

Finally, I am very grateful for the support of my family and friends who encouraged and supported me throughout my years of study and through the process of researching and writing this thesis. Especially I want to thank Konrad, my better half, who always tried to keep me happy and made everything appear easier.

Abstract (en)

The recent debate questioning validity and reliability of functional magnetic resonance imaging (fMRI) results illustrates the importance of the present topic. Especially in the clinical context, methodological assumptions which have been developed for fMRI analyses of healthy subjects, may not be suitable. Much more, reliable single subject analyses are needed in the clinical context. One approach for this purpose is the risk map technique. It is a method with minimized methodological assumptions to stay close to the original data. Validity is tried to be improved by extracting the most reliably activated voxels based on repeatability. We applied the risk map technique on brainstem scans to find out if reliable results can be yielded in this area. The brainstem, however, is a very challenging part of the brain for fMRI scans, due to an increased level of physiological noise and its anatomical peculiarities.

Therefore, we first tried to get an overall estimation of the blood oxygenation level-dependend (BOLD) signal strength in the brainstem motor nuclei as compared to the motor cortex, by comparing detectable activations of a motor task in motor cortex and brainstem trigeminal motor nuclei. Here we found activated voxels in the cortex in 100% of the cases, whereas only 36% were detected in brainstem. Further, brainstem activations were found with lower correlation thresholds than cortex activations. These results provide information about how challenging fMRI scans in the brainstem motor nuclei are, compared to the motor cortex.

In the next steps, we extracted measures of brainstem motor nuclei activation, to find out if and how data quality may be improved. We used four different preprocessing conditions. One without smoothing, one where data have been smoothed, and two conditions with smoothing and masking. In one case cerebrospinal fluid (CSF) masks, in the other case brainstem (BS) masks have been drawn. From the risk map analyses, we found that our data benefit from spatial smoothing, with and without masking. In conditions with smoothed data, more activation could be detected. Therefore, smoothing reduces some of the noise in brainstem scans.

As an additional method to evaluate the results of the risk map technique, receiver operating characteristic (ROC) curve analysis has been used. This method gives information about the sensitivity and specificity of our test conditions. Here we could show that sensitivity and specificity could be improved with masking conditions, probably because artefacts are reduced in advance.

Finally, we tested the influence of different reference functions of the hemodynamic response on our data. We found that the function “Shift +1”, a step function with 5 seconds (2 TR) latency, led to best results in risk map technique and ROC curve analysis. This finding indicates a delayed hemodynamic response in brainstem motor nuclei.

Even though we could show an improvement of BOLD signal strength of our data by spatial smoothing and improved sensitivity and specificity with masking conditions, results indicate that with the number of repeated scans given in our investigation, voxel reliability detected with standard risk maps is much below the performance in cortical areas. Whereas the technique is well suited for clinical diagnostics of cortical brain activations, using the risk map technique for brainstem fMRI is very challenging. The scans come with an increased level of noise which is problematic for all analysis techniques. Comparable to the electrophysiological evoked response recordings, clinical application of an fMRI analysis technique with minimized methodological assumptions would require more scan repetitions – a limiting factor for a patient investigation. The data also point out that – irrespective of the analysis technique used - in order to improve validity of the functional brain stem results, it is important to avoid as many artifacts as possible already during the data recording process.

Abstract (de)

Die derzeit herrschende Debatte, welche die Validität und Reliabilität funktioneller Magnetresonanztomographie (fMRT) Ergebnisse in Frage stellt, verdeutlicht die Wichtigkeit des vorliegenden Themas. Vor allem im klinischen Kontext sind methodische Annahmen, die für fMRT-Auswertungen gesunder Versuchspersonen entwickelt wurden, möglicherweise nicht passend. Viel eher werden reliable individual-Analysen im klinischen Kontext gebraucht. Daher müssen alternative Methoden berücksichtigt werden. Ein Ansatz für diesen Zweck ist die Risk Map Technik. Es ist eine Methode mit minimierten methodischen Annahmen, um nahe an den ursprünglichen Daten zu bleiben. Es wird versucht die Validität zu verbessern, indem die zuverlässigsten Voxel, basierend auf Wiederholbarkeit, extrahiert werden. Wir haben die Risk Map Technik auf Hirnstamm-Scans angewendet um herauszufinden, ob zuverlässige Ergebnisse in diesem Areal erzielt werden können. Es ist jedoch schwierig fMRT-Auswertungen im Hirnstamm durchzuführen, wegen des erhöhten Niveaus physiologischen Rauschens und seiner anatomischen Besonderheiten.

Daher wollten wir zunächst eine Gesamtschätzung der blood oxygenation level-dependend (BOLD)-Signalstärke in den Hirnstamm-Kernen, im Vergleich zum motorischen Kortex bekommen, indem wir detektierbare Aktivitäten einer motorischen Aufgabe im motorischen Kortex mit der in trigeminalen motorischen Hirnstammkernen verglichen. Wir fanden aktivierte Voxel im Kortex in 100% der Fälle, im Hirnstamm jedoch nur in 36% der Fälle. Außerdem wurde die Hirnstammaktivität bei niedrigeren Korrelationsschwellen gefunden als die Kortexaktivität. Diese Ergebnisse gaben uns Informationen darüber, wie herausfordernd fMRT-Scans in motorischen Hirnstamm-Kernen im Vergleich zum motorischen Kortex sind.

In den nächsten Schritten führten wir Messungen der Aktivierung motorischer Hirnstammkerne durch, um herauszufinden, ob und wie die Datenqualität verbessert werden kann. Wir verwendeten vier verschiedene Vorverarbeitungsbedingungen. Eine ohne Glättung, eine bei der die Daten geglättet wurden, und zwei Bedingungen mit Glättung und Maskierung. In einem Fall wurden Cerebrospinalflüssigkeits (CSF)-Masken, im anderen Fall Hirnstamm (BS)-Masken verwendet. Unsere Risk Map Analyse zeigte eine Verbesserung der Ergebnisse durch räumliche Glättung, sowohl mit, als auch ohne Maskierung. In Bedingungen mit geglätteten Daten konnte mehr Aktivität festgestellt werden. Das glätten scheint daher einen Teil des Rauschens der Hirnstamm-Scans zu reduzieren.

Als zusätzliche Methode um die Resultate der Risk Map Technik zu evaluieren, wurde die Receiver-Operating-Characteristic (ROC)-Kurve verwendet. Mit dieser Methode gewinnen wir

Informationen über die Sensitivität und Spezifität unserer Tests. Hier konnten wir zeigen, dass unsere Maskierungsbedingungen die Sensitivität und Spezifität verbesserten, wahrscheinlich weil dadurch Artefakte im Voraus reduziert wurden.

Zuletzt testeten wir den Einfluss verschiedener Referenzfunktionen der hämodynamischen Reaktion auf unsere Daten. Wir haben festgestellt, dass die Funktion "Shift +1", eine Step-Funktion mit 5 Sekunden (2 TR) Latenzzeit, zu den besten Ergebnissen bei der Risk Map Technik und der ROC Kurvenanalyse führte. Dieses Ergebnis deutet auf eine verzögerte hämodynamische Reaktion in motorischen Hirnstamm-Kernen hin.

Obwohl wir eine Verbesserung der BOLD-Signalstärke unserer Daten durch räumliche Glättung, sowie verbesserte Sensitivität und Spezifität mit Maskierungs-Bedingungen zeigen konnten, weisen die Ergebnisse darauf hin, dass mit der Anzahl der in unserer Untersuchung wiederholten Scans, die Reliabilität der Voxel, die mittels Risk Map Technik nachgewiesen wurde, erheblich unter der Leistung in kortikalen Bereichen liegt. Während die Technik gut für klinische Diagnostik kortikaler Hirnaktivierungen geeignet ist, ist die Verwendung der Risk Map Technik für Hirnstamm-fMRT eine große Herausforderung. Die Scans weisen erhöhtes Rauschen auf, was für jede Analysetechnik problematisch ist. Vergleichbar mit Aufzeichnungen über elektrophysiologische evozierte Potentiale, würde die klinische Anwendung einer fMRT-Analysetechnik mit minimierten methodischen Annahmen mehr Scan-Wiederholungen erfordern – ein limitierender Faktor bei Patientenuntersuchungen. Außerdem weisen die Daten darauf hin, dass es - unabhängig von verwendeter Analysetechnik - zur Verbesserung der Validität der funktionellen Hirnstamm-Ergebnisse wichtig ist, bereits bei der Datenerfassung möglichst viele Artefakte zu vermeiden.

Table of Contents

1. Introduction

A major topic in functional magnetic resonance imaging (fMRI) is the validity and reliability of results. In 2012 and 2016 Eklund et al. pointed out the risk for obtaining false positive fMRI results, when using some of the common fMRI software packages (SPM, FSL, AFNI) and analyses. They stated that both, group analyses and single-subject analyses are affected by this problem and that it probably arises from imperfect methodological assumptions (Eklund et al., 2016, 2012). Particularly for the clinical context methodological assumptions which have been developed for healthy subjects may not be suitable. Therefore other methods to obtain valid results need to be considered. The risk map technique, developed in the context of clinical fMRI, is an attempt to yield valid fMRI results using minimized methodological assumptions. The idea is to improve validity of fMRI scans by extracting most reliably activated voxels based on their repeatability. It has been shown that risk map analysis can improve validity of presurgical evaluation in cortex areas and may successfully be used for clinical diagnostics (Beisteiner, 2017; Beisteiner et al., 2000).

Ogawa et al. (1990) introduced the blood oxygenation level-dependend (BOLD) effect as the basis for the contemporarily most common fMRI method. It makes use of the fact, that during a hemodynamic response (response of vascular system to neuronal activity), the level of oxygenated hemoglobin (HbO_2) is raised, as the regional cerebral blood flow (rCBF) is increased. Thus, the ratio of oxygenated (HbO_2) to deoxygenated (Hb) hemoglobin increases. HbO_2 is diamagnetic and does not strongly affect the local magnetic field. Hb, on the other hand, is paramagnetic and alters the local magnetic susceptibility, leading to magnetic field distortions in the surrounding capillary bed and venules. Therefore, the increased level of HbO_2 during the hemodynamic response leads to a more homogeneous local magnetic field and thus, to a stronger MRI signal. The BOLD effect, therefore, indirectly measures increased neuronal activity, by detecting the local magnetic field homogeneity, caused by an increase of HbO_2. (Ogawa et al., 1990; Stippich, 2015)

The brainstem is the evolutionary oldest part of mammalian brains and most important for regulation of vital functions like respiration, consciousness and body temperature. It is densely packed with nuclei and serves as relay between cerebrum, spinal cord and cerebellum (Bear et al., 2007; Beissner et al., 2011). Unfortunately, the brainstem is also target in various diseases, such as Parkinson's disease, spinocerebellar ataxia and various headache disorders (Beissner, 2015; Braak et al., 2004; Schulte et al., 2016). However, not many fMRI studies addressing brainstem functions were conducted so far, probably due to the challenging

increased level of physiological noise. In particular, up to now only one brainstem fMRI study on single-subject level exists (Matt et al., in submission), as opposed to group studies. The brainstem lies in close proximity to large blood vessels and the cerebrospinal fluid (CSF), leading to strong motion and physiological artifacts. Further, brainstem nuclei are smaller than cortical structures that are commonly investigated with fMRI, and many nuclei with different functions lie closely adjacent (Beissner, 2015; Beissner et al., 2014; Schulte et al., 2016).

Several suggestions were made to improve this challenging situation. To cope with the small size of brainstem nuclei, an increase in spatial resolution is favorable. As this reduces the BOLD signal strength, more volumes must be acquired or higher field strengths can be used to increase the sensitivity. The strong physiological noise can be reduced by correcting data with physiological signals (Beissner, 2015; Matt et al., in submission). Further, Beissner et al. (2014) suggested to use anatomically-defined masks, cropped to the brainstem, to exclude adjacent structures. With this approach, noise or signals that arise from surrounding subcortical and cerebellar structures, can be excluded. Matt et al. (in submission) additionally used CSF masks, where voxels were probabilistically assigned to tissue classes like the CSF, which then could be excluded from further analyses. Using this approach, motion artifacts and physiological artifacts, arising from the CSF can be reduced.

Being able to use functional imaging in the brainstem could be a major advantage in the clinical context, as this part of the brain is affected by many diseases. However, obtaining valid results despite the increased physiological noise seems very challenging. In this thesis we want to find out whether – besides of the proven efficiency for cortical activations - it is possible to detect valid and reliable BOLD signals also from the brainstem, with a standard version of the clinical risk map technique. An important issue for clinical applications is, that they require analyses of the data on single-subject level.

In order to answer our question, we first compared detectable activations of a motor task in motor cortex and brainstem trigeminal motor nuclei, using the standard risk map technique. This first approach should help us to get an overall estimation of the BOLD signal strength in brainstem nuclei as compared to motor cortex, when they are analyzed with the risk map technique.

In the next steps, solely brainstem BOLD signals were addressed. We extracted measures of brainstem trigeminal motor nuclei activation to find out if and how data quality can be improved. Therefore, we analyzed our brainstem data with four different preprocessing conditions to determine which of the conditions may improve data quality. In the first condition, data were not smoothed and no masks were used. For the second condition, images were smoothed, but not

masked. In condition three and four, images were smoothed and masks were used. Cerebrospinal fluid (CSF) masks were used in condition three, brainstem (BS) masks in condition four. A detailed description of the preprocessing conditions and masks is given in 2.4.

Finally, we investigated three different reference functions of the hemodynamic response. It has been shown that the hemodynamic response can be altered by pathologies like brain tumors (Wang et al., 2012) or arteriosclerotic changes (D'Esposito et al., 2003). Further, Handwerker et al. (2004) showed variations of the hemodynamic response across healthy subjects and brain regions. Therefore, we analyzed our results with three different reference functions to investigate which one is most suitable for trigeminal motor nuclei activation in the brainstem.

2. Methods

2.1. Participants

6 healthy subjects (3 female) took part in the study. Their mean age was 29.7 years. All subjects gave written informed consent according to the Declaration of Helsinki. The ethics committee of the Medical University of Vienna approved the study.

2.2. Task description

The participants performed a jaw clenching task, as it has already been shown by Beissner et al., (2011) that this kind of motor task can activate the trigeminal motor nuclei. The participants were asked to repeatedly clench their teeth once every second, while keeping their mouth closed, in order to reduce task-related motion.

Each subject participated in 4 repeated fMRI sessions. The median interval between the measurements were 13 days. As shown in Figure 1, a block design was used, consisting of 3 blocks of motor task and 4 blocks of rest (in each case 20 seconds). 8 runs have been performed in each session. For further details please see Matt et al. (in submission).

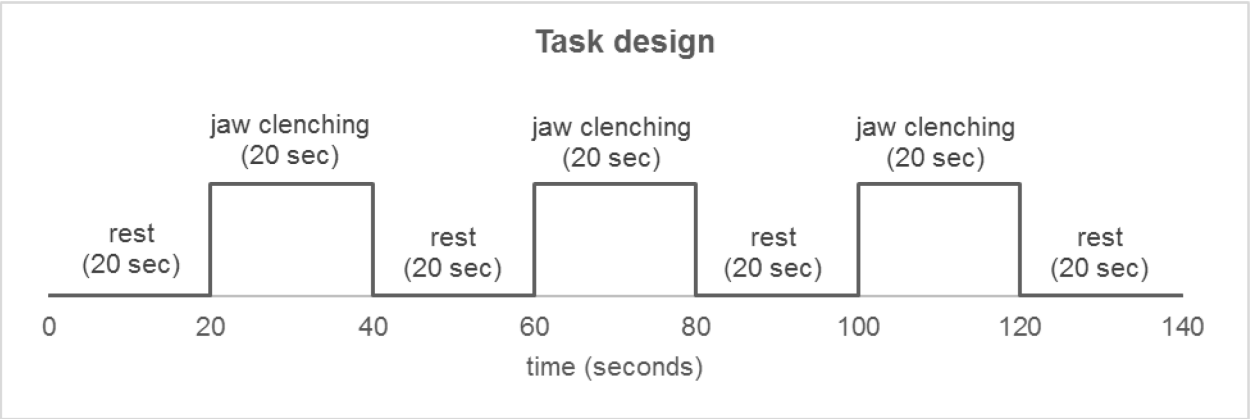

Figure 1. Task design of fMRI sessions. Block design with 20-sec-phases of rest (4 blocks) and task (3 blocks). 8 runs were done in each scanning session.

2.3. Measurement parameters

Images were acquired using a 7 T MAGNETOM system (Siemens Healthcare, Erlangen, Germany), with a 32 channel Nova Medical head coil (Wilmington, USA). An optimized multiband EPI sequence was used, with 30 coronal slices oriented parallel to the floor of the fourth ventricle. 140 volumes per run were acquired. The parameters were: TR/TE=1000/23 ms, multiband factor = 2, in-plane acceleration = GRAPPA 2, echo spacing = 0.81 ms, bandwidth = 1450 Hz/Px, FOV: 220 x 220 mm, in-plane resolution: 1.46 x 1.46 mm, slice thickness: 1.5 mm, gap: 15%, with a foot-to-head phase-encoding direction. Prior to the functional scans, B0 field maps were acquired. The same slice prescription as the EPI sequence was used, with a multi‑echo gradient-echo (MGE) sequence, with the parameters TR = 800 ms, TE = 5, 10, 16 ms.

2.4. Image preprocessing

Image preprocessing was carried out using SPM12 (Wellcome Trust Centre for Neuroimaging; www.fil.ion.ucl.ac.uk/spm/software/spm12). First, images of all sessions per subject were realigned to the mean volume. Then a distortion correction using the session-specific field map (Robinson and Jovicich, 2011) was done. Afterwards the images were coregistered.

For further comparisons, 4 different preprocessing conditions were used. For condition 1 “No Mask, No Smoothing” the images were not further processed. For condition 2 “No Mask, Smoothing”, images were smoothed (3 mm FWHM Gaussian kernel). Condition 3 “CSF Mask,

Smoothing" was gained by first creating individual cerebrospinal fluid (CSF) masks, where probabilistically assigned CSF-voxels were excluded for further analyses (see Figure 2). CSF masks were acquired by SPM segmentation of the coregistered mean EPI, with a threshold at 0.25. Afterwards images were smoothed (like in condition 2). For condition 4 "BS Mask, Smoothing" individual brainstem (BS) masks were created. BS masks were drawn on the individual coregistered mean EPI using MRIcron and include medulla, pons and mesencephalon (see Figure 3). In this condition only the area of the BS mask is used for further analyses. After creating the masks, the images were also smoothed (like in condition 2). (Matt et al., in submission)

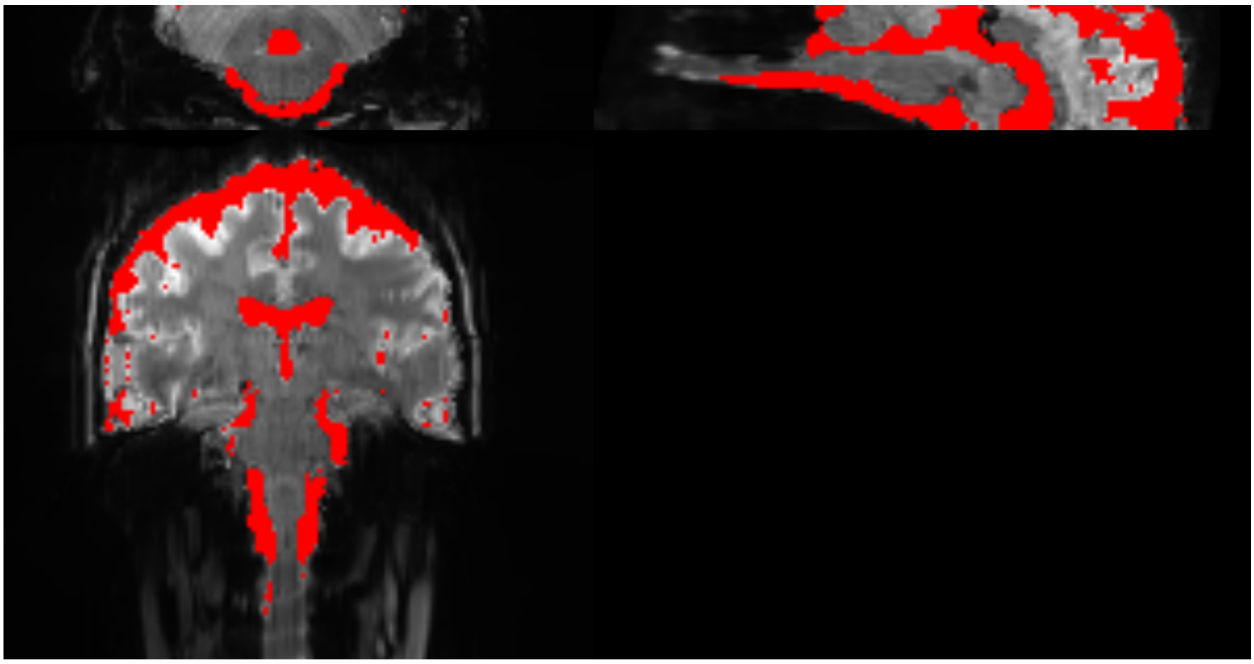

Figure 2. Cerebrospinal fluid (CSF) mask. Example for CSF mask drawn on subject BS02, acquired by SPM segmentation of the coregistered mean EPI, with a threshold at 0.25. Shown are axial (top left image: top is posterior), coronal (bottom left image: left side is right hemisphere) and sagittal sections (top right image: right is top of the brain, left is myelon). CSF Mask, comprising probabilistically assigned CSF-voxels, is excluded for further analyses.

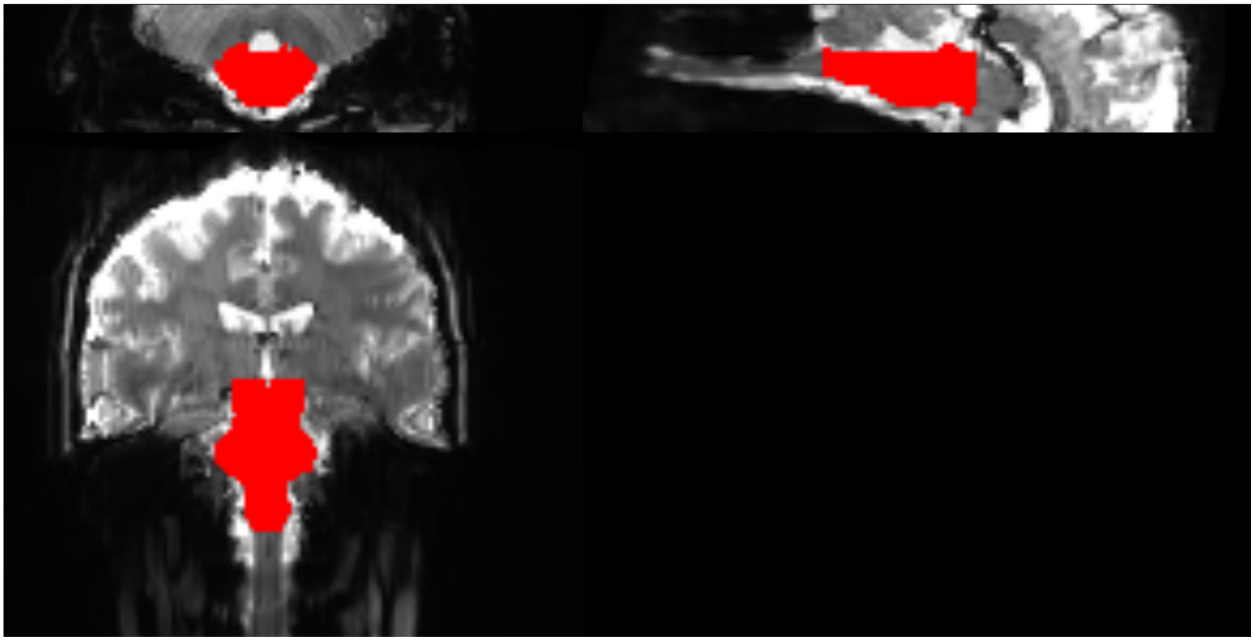

Figure 3. Brainstem (BS) mask. Example for BS mask drawn on subject BS02, using the individual coregistered mean EPI in MRIcron. Shown are axial (top left image: top is posterior), coronal (bottom left image: left side is right hemisphere) and sagittal sections (top right image: right is top of the brain, left is myelon). BS mask includes medulla, pons and mesencephalon and is the only area included for further analyses in this condition.

2.5. Data analysis

2.5.1. Regions of interest (ROI)

The target ROIs assessing the trigeminal motor nuclei were drawn bilaterally and for each subject separately, using the mean EPI in MRICron. The exact anatomical location was based on the coordinates published by Beissner et al. (2011). As control ROI, an area in the high pons was chosen. This area doesn't contain motor or sensory nuclei. The location was based on the Duvernoy's Atlas of the Human Brain Stem and Cerebellum (Naidich et al. 2009). In Figure 4 bilateral target ROIs in red and control ROIs in green are shown.

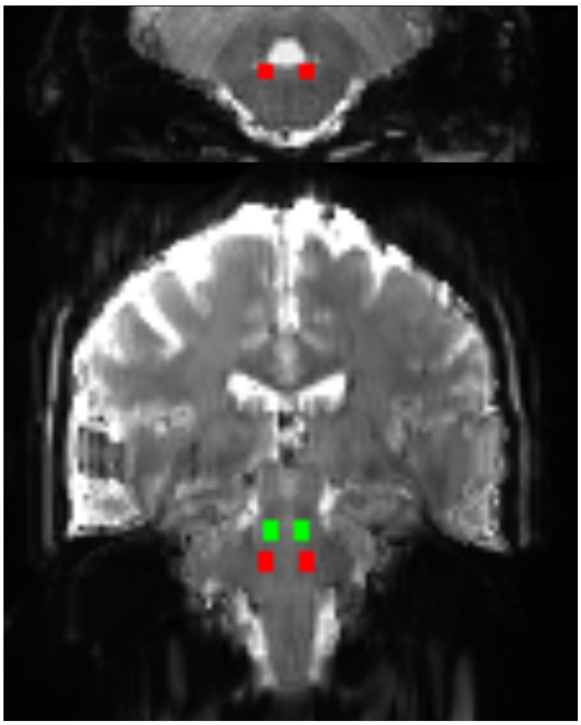

Figure 4. Regions of interest (ROI). This figure shows target ROIs (red squares) and control ROIs (green squares), both of which comprise 72 voxels (3 x 3 x 4 voxels per side), of subject BS04, in axial (top image: top is posterior) and coronal (left side is right hemisphere) sections. Target ROIs cover trigeminal motor nuclei and were drawn bilaterally and for each subject separately, using the mean EPI in MRICron. Control ROIs are located in high pons and do not contain motor or sensory nuclei.

Each ROI has a size of 3 x 3 x 4 voxels per side, which is roughly 4.5 x 4.5 x 6 mm. Therefore the bilateral ROIs comprise 72 voxels (~240 mm^3). (Matt et al., in submission)

ROIs have not been used to identify activation in motor cortex, as BOLD signal strength in this area is very high and it has repeatedly been shown that assessing motor cortex activation with risk maps works well. (Beisteiner et al., 2000; Roessler, 2005). Instead, the location of the motor cortex region responsible for jaw movement, was defined neuroanatomically. The area was selected bilaterally and for each subject separately, using the mean EPI in MRICron (see Figure 5).

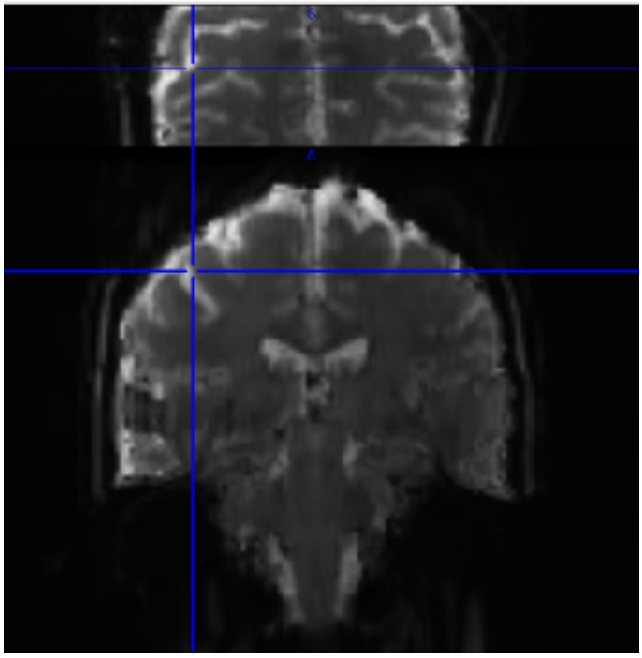

Figure 5. Motor cortex area for jaw movement. The images show the area for jaw movement of the primary motor cortex, marked with the blue cross, in axial (top image: top is posterior) and coronal (left side is right hemisphere) sections of subject BS02. The area was defined neuroanatomically, using the mean EPI in MRICron.

2.5.2. Risk Map Technique:

The Risk Map technique was developed for presurgical fMRI. The idea is to extract the most reliably activated voxels of multiple measurements. It is a technique with minimized methodological assumptions, which is an advantage in the context of presurgical fMRI, since the variability and latency of BOLD signal courses in pathological brains cannot be estimated and multiple model assumptions with standard postprocessing techniques bear a high risk for false positive or false negative results in individual patients. In addition, risk maps do not mask artifacts and therefore allow an immediate judgement of scan / patient quality and general signal validity. (Beisteiner, 2017; Beisteiner et al., 2000)

Risk maps are created first, by calculating Pearson correlation *r* between the BOLD signal time course and the reference functions (described in section 2.5.3) for each voxel and run. Second, the voxels are color coded by their repeatability of activation across all runs, at a given correlation threshold. The color code is the following: a voxel is colored yellow when it is above threshold in >75% of all runs, orange in >50% and a red voxel is above threshold in >25% of all

runs. This last step, the color coding, is repeated for various thresholds and as such depends on the respective preselected correlation threshold. (Beisteiner, 2017; Beisteiner et al., 2000)

The calculated risk map pictures were then assembled, as shown in Figure 6. Per subject, one picture-block was created, for each preprocessing condition of each session. Figure 6 shows risk map pictures of subject BS05, session 2. Each of the picture-blocks contains the risk map pictures for the three different reference functions used, and for the correlation thresholds from 0.2 to 0.95. The red rectangle of Figure 6 is shown enlarged in Figure 7. Each square in Figure 7 contains the risk map pictures of all brain slices that have been acquired, each time for a specific correlation threshold. For the evaluation of the pictures, the highest correlation threshold still containing colored voxels within the ROIs was searched.

Risk maps are evaluated by visual inspection. Our data resulted in a huge number of pictures needed to be evaluated. Additionally, brainstem scans come along with many artefacts in the images, resulting in noisy risk maps. In order to minimize biases resulting from visually evaluating a lot of noisy pictures, ROIs have been used. As already mentioned, no activation and therefore no colored voxels should occur in our control ROIs. If colored voxels were observed in the control ROIs, the picture was dismissed from defining functional activation (see Figure 8). Therefore, trigeminal motor nuclei were only counted as activated, in case of at least one colored voxel in one or both target ROIs and none colored voxel in both control ROIs (see Figure 9).

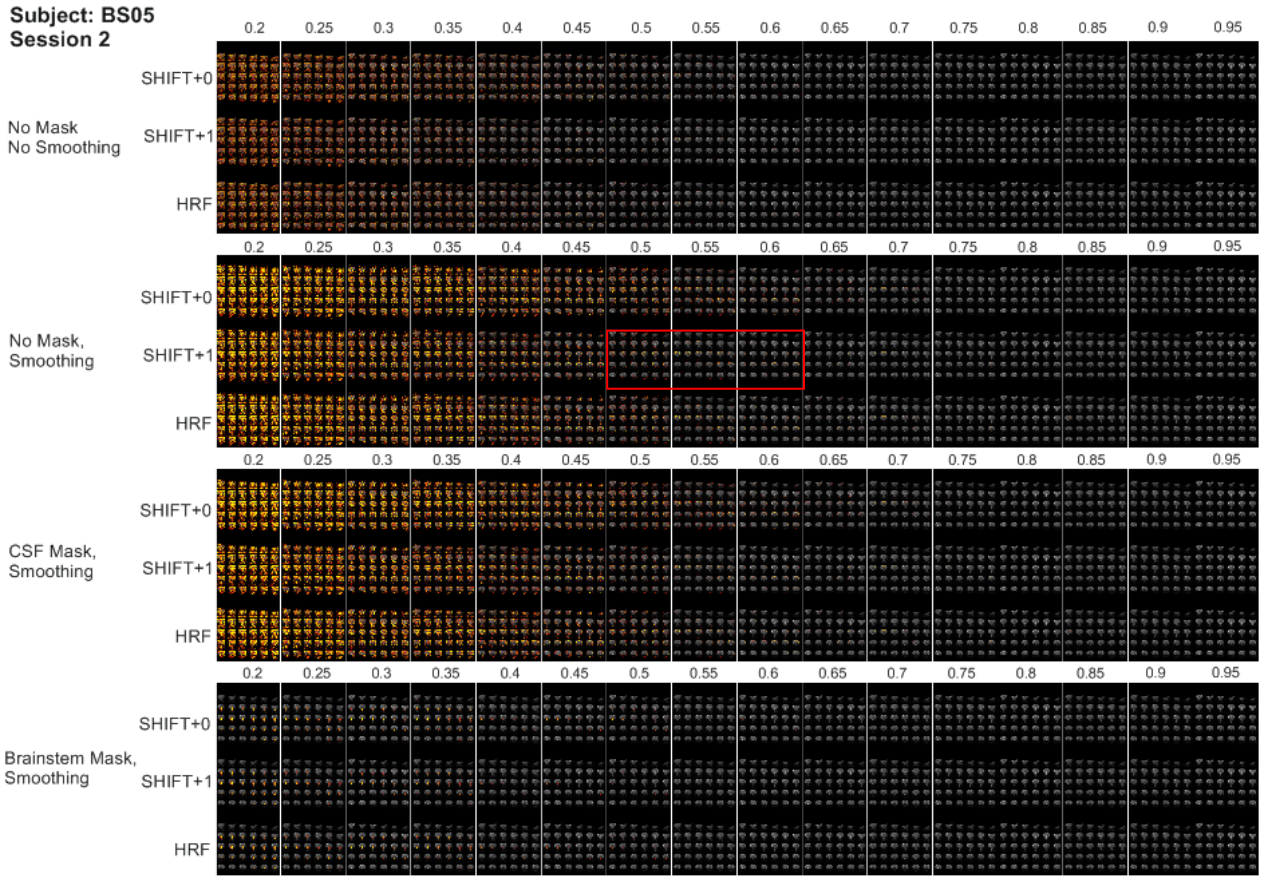

Figure 6. Assembled risk map pictures of subject BS05, Session 2. One picture-block is made for each preprocessing condition. Each picture-block consists of the three different reference functions. The part with the red rectangle is shown enlarged in Figure 7.

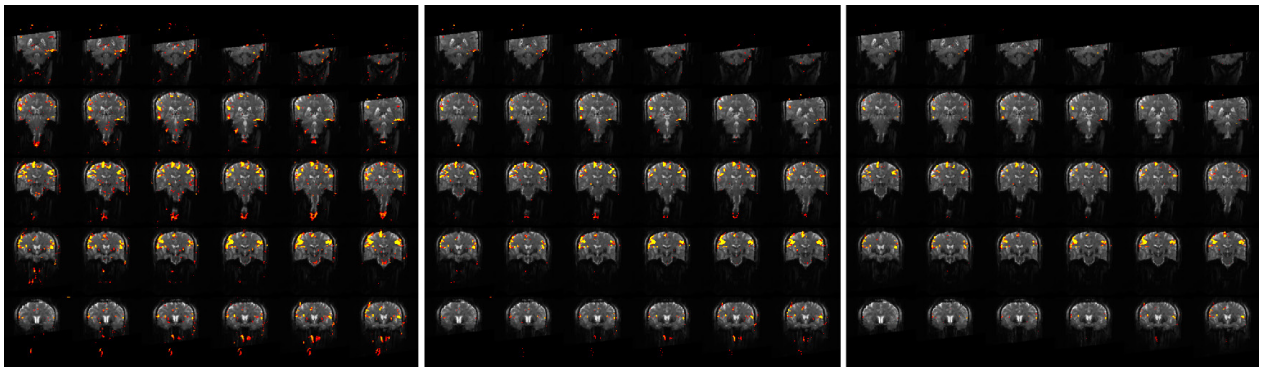

Figure 7. Enlarged part of assembled risk map pictures. Subject BS05, Session 2, condition "No Mask, Smoothing", reference function "Shift +1", correlation thresholds 0.5, 0.55, 0.6. Each square consists of the acquired brain slices for a specific correlation threshold.

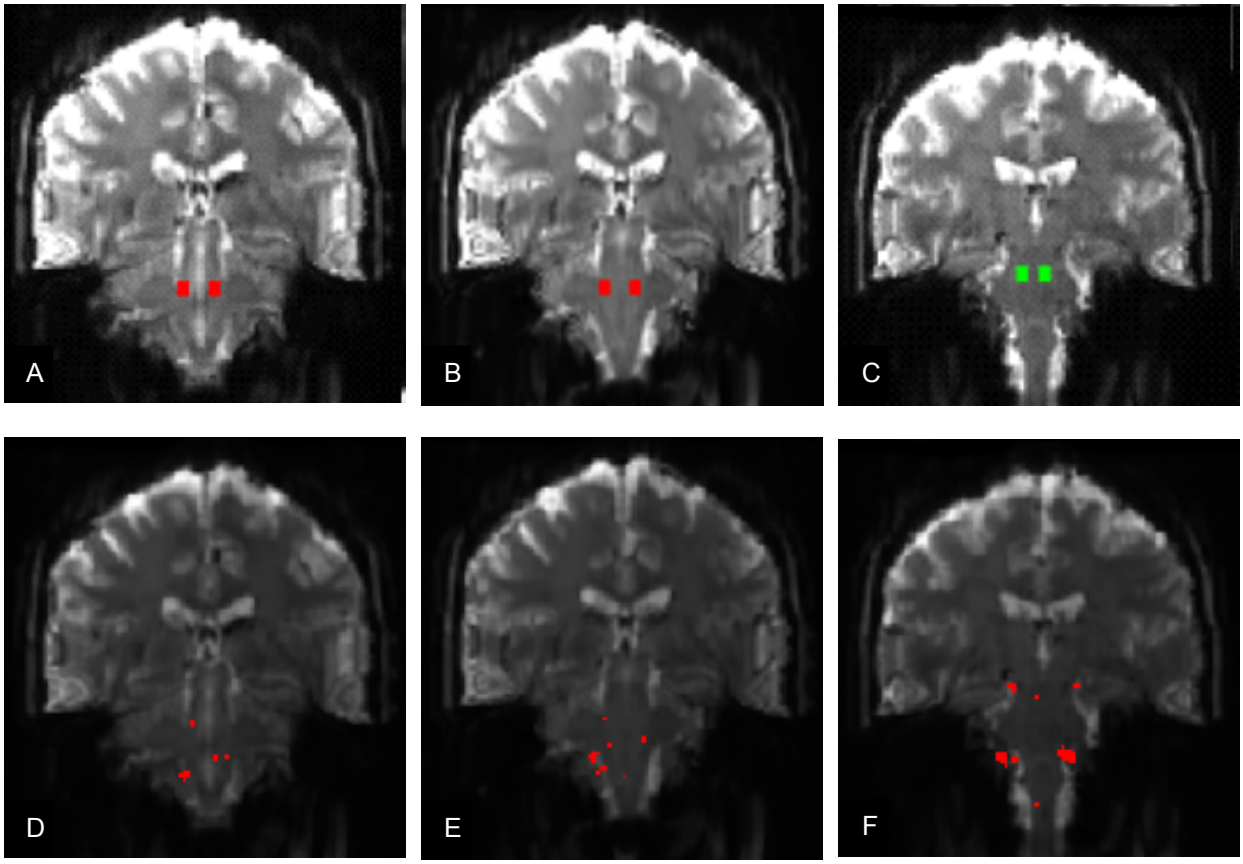

Figure 8. A-F Example for activation in both, target ROI and control ROI. Subject BS02, session 4, condition “BS Mask, Smoothing”, reference function “Shift +1”, correlation threshold 0.35. Shown are coronal (left is right hemisphere) sections. The three images in each row show three different slices of the same person and condition set. **A-C** Reference pictures with target ROIs (red squares, see **A, B**) in which we would expect activation. Control ROI (green squares, see **C**) in which we would not expect activation. **D-F** Calculated risk maps. **D** Same slice as **A**, activation within target ROI (right hemisphere). **E** Same slice as **B**, activation within target ROI (right hemisphere). **F** Same slice as **C**, colored voxel within control ROI (right hemisphere).

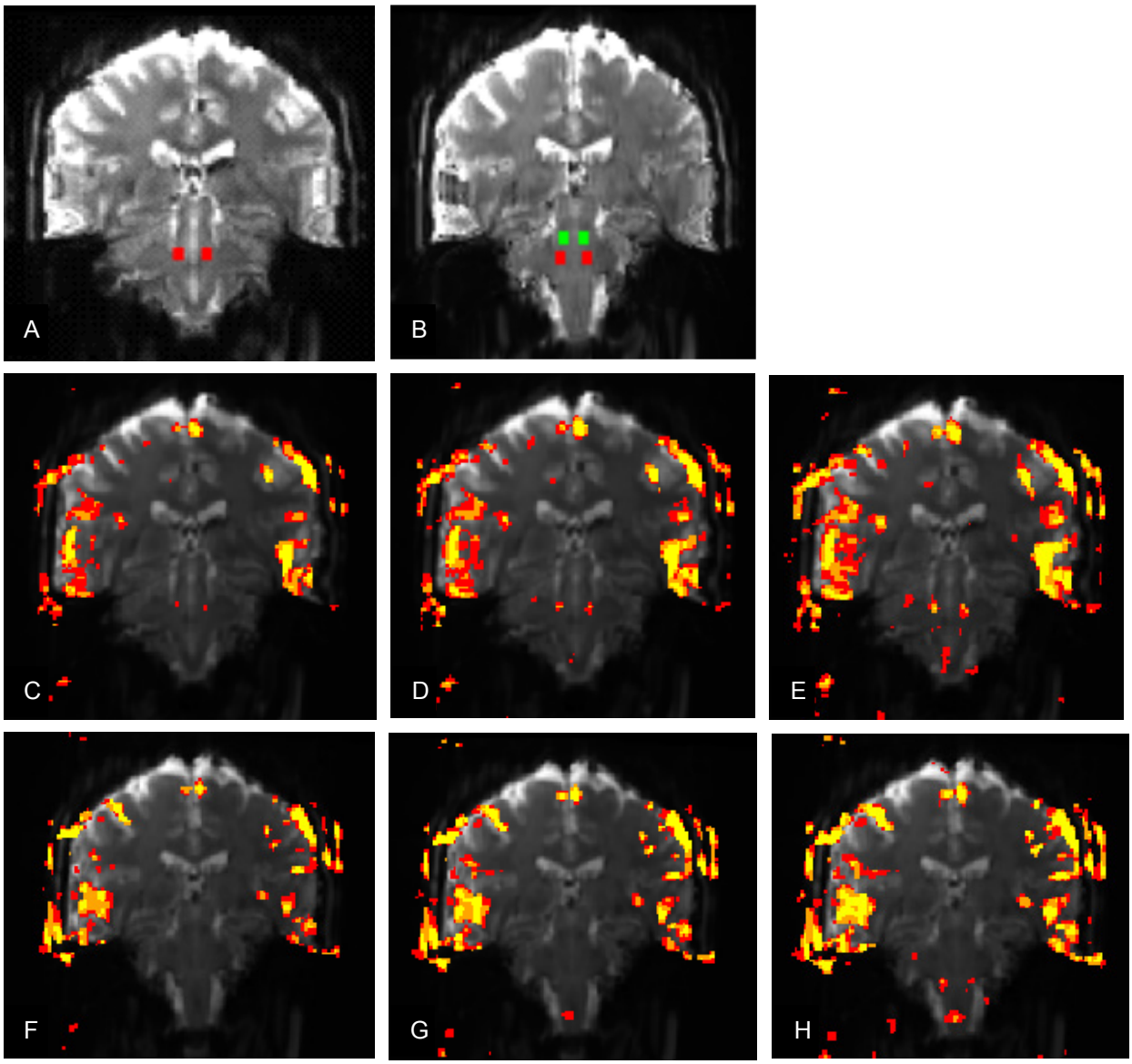

Figure 9. A-H Example for activation in target ROI without activation in control ROI. Subject BS02, session 1, condition “CSF Mask, Smoothing”, reference function “Shift +0”. Shown are coronal (left is right hemisphere) sections. **A, B** Reference pictures with target ROIs (red squares, see **A, B**) in which we would expect activation and control ROIs (green squares, see **B**) in which we would not expect activation. The two images show two different slices of the same person and condition set. **C-E** Same slice as **A**, at correlation thresholds 0.4 (**C**), 0.35 (**D**), 0.3 (**E**). Activation within target ROIs starts at 0.4 and expands with decreasing correlation threshold. **F-H** Same slice as **B**, at correlation thresholds 0.4 (**F**), 0.35 (**G**), 0.3 (**H**). No colored voxels occur within control ROIs.

2.5.3. Reference Functions

The classical approach to model the BOLD hemodynamic response is the Hemodynamic Response Function (HRF). A schematic representation of a HRF for one stimulus is shown in Figure 10. The assumption is, that the BOLD signal rises after 2-3 second and reaches a peak after 5-6 seconds. Baseline level is reached after about another 10 seconds (Stippich, 2015). As we used a block design, the HRF will remain on a plateau phase in our case. However, it has been shown that HRFs can vary across subjects and brain regions, which may have an impact on results of statistical analyses (Handwerker et al., 2004). Of particular importance are possible HRF variations in the context of clinical fMRI, as pathologies e.g. brain tumors can alter the hemodynamic response (Beisteiner, 2017;). Even though we assessed healthy subjects, additional reference functions were used to consider variabilities of the hemodynamic response in the brainstem. The additional reference functions are step functions with 2.5 seconds latency (referred to as “Shift +0”, shown in Figure 11) and with 5 seconds latency (referred to as “Shift +1”, shown in Figure 12). (Beisteiner, 2017)

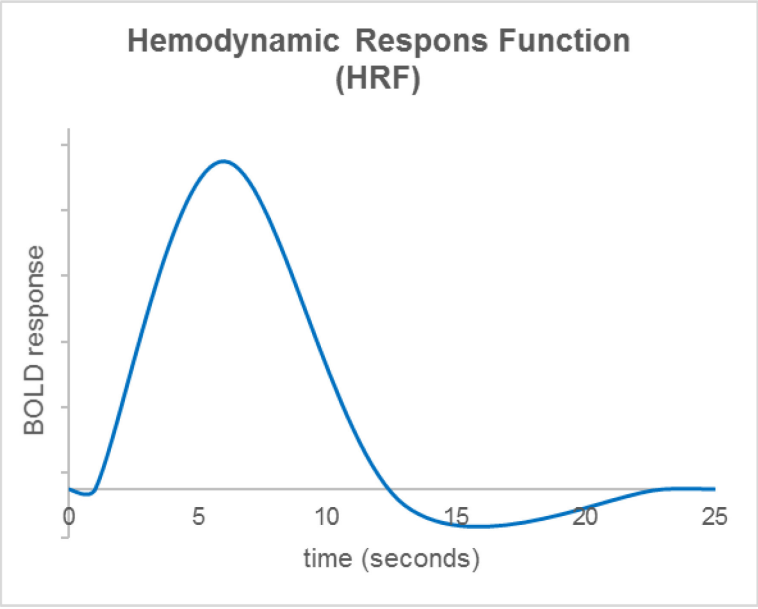

Figure 10. Example of hemodynamic response function (HRF). BOLD signal rises after 2-3 second and reaches a peak after 5-6 seconds. Baseline level is reached after about another 10 seconds.

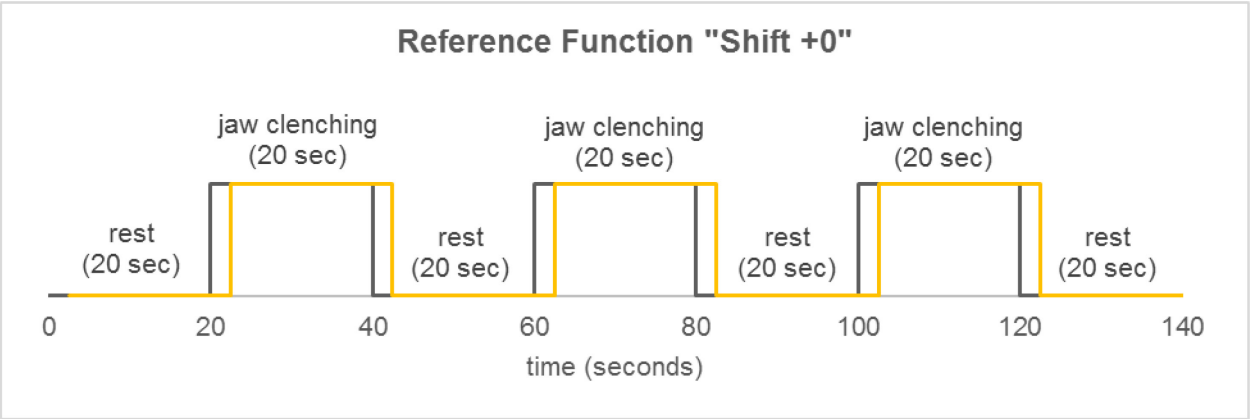

Figure 11. Reference function "Shift +0". Step function with 2.5 sec. latency. "Shift +0" (yellow line) is shifted 2.5 sec. in respect to the designed task (black line).

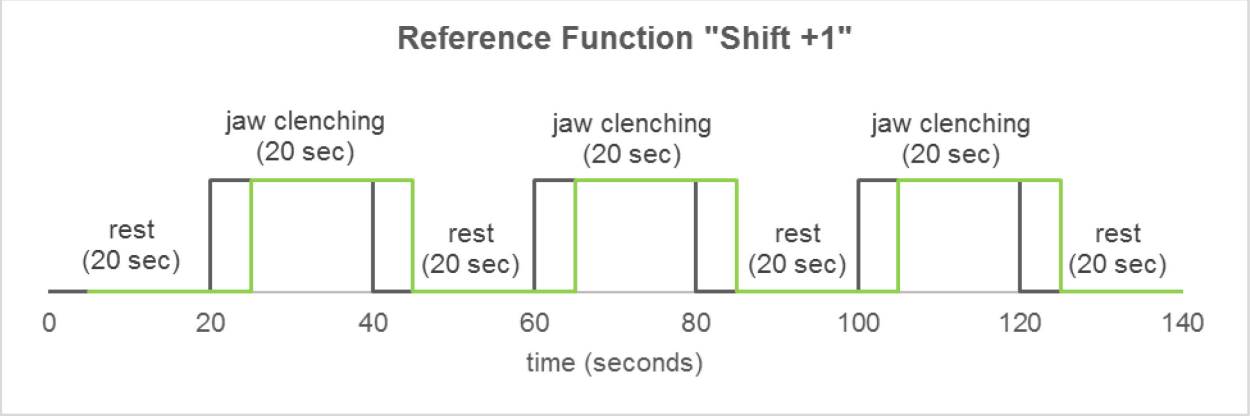

Figure 12. Reference function "Shift +1". Step function with 5 seconds latency. "Shift +1" (green line) is shifted 5 sec. in respect to the designed task (black line).

2.5.4. Receiver Operating Characteristic (ROC) Curve Analysis:

Receiver operating characteristic (ROC) curves are widely used to evaluate the quality of diagnostic tests. A ROC curve is created by plotting sensitivity (true positive rate (TRP)) on the y-axis against 1-specificity (false positive rate (FPR)) on the x-axis. In our case, correlation maps from the risk map analysis have been z-transformed and used as basis for calculating ROC curves, where positive values in the target ROI were defined as true positive values, and positive values in the control ROI as false positive values (see Table 1) (Grzybowski and Younger, 1997; Matt et al., in submission; Park et al., 2004).

Table 1. Overview of classification of values and calculations for receiver operating characteristic (ROC) curves.

	ROI	
Z-value	Target ROI	Control ROI
Positive	TP	FP
Negative	FN	TN

True positive rate (TPR) = Sensitivity = TP/(TP+FN)

True negative rate (TNR) = Specificity = TN/(FP+TN)

False positive rate (FPR) = 1-Specificity

These calculations are done for all z values detected in the ROIs. In our case, all ROC curve figures contain 3 curves, one for each reference function. Figure 13 shows an example of our ROC curves with the different reference functions.

A common way to quantify the performance of a test with ROC curve analysis, is the Area under curve (AUC). Its value can range between 0 and 1. A 45-degree reference line is plotted, which has an AUC of 0.5, corresponding to no discriminatory power of the test (see Figure 13). The higher the AUC value, the better the sensitivity and specificity and therefore the overall accuracy of the test. (Grzybowski and Younger, 1997; Park et al., 2004)

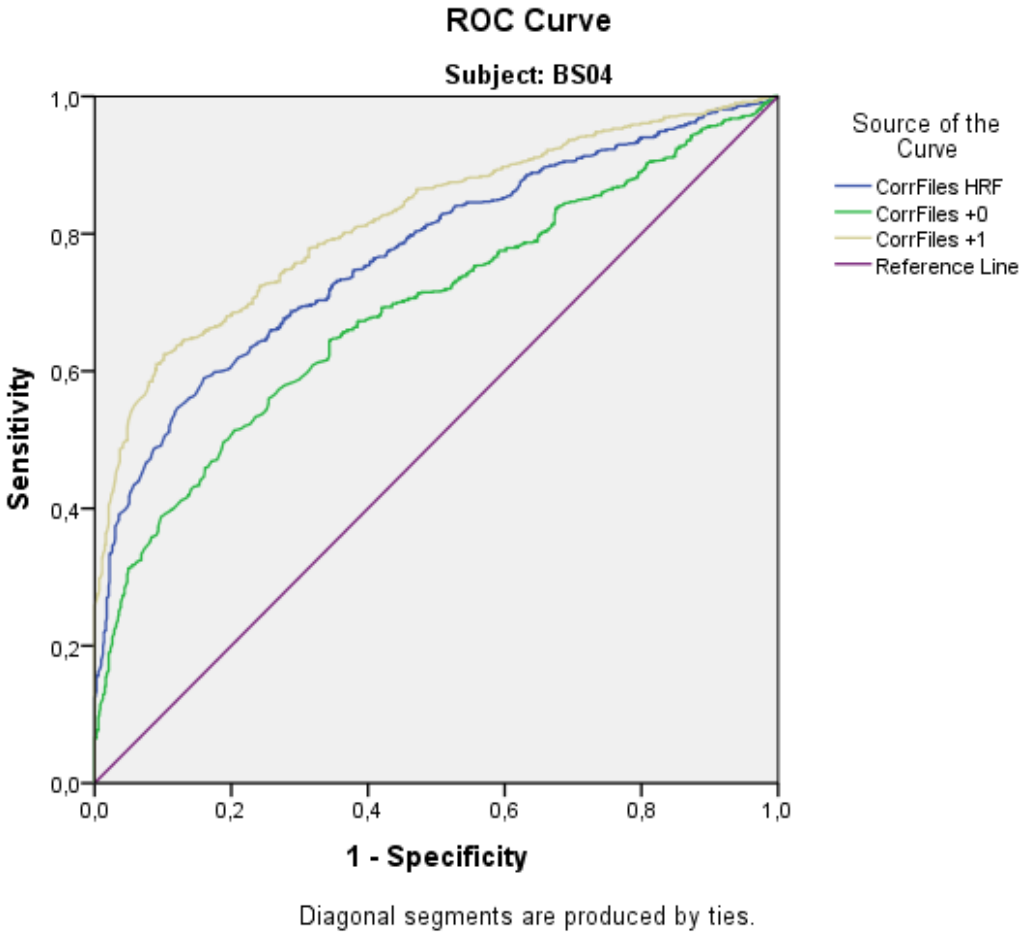

Figure 13. Example of receiver operating characteristic (ROC) curve. This figure shows a calculated ROC curve of subject BS04. In each of our ROC curves, three lines are shown, representing the three reference functions “HRF”, “Shift+0” and “Shift +1”. Additionally, a 45° reference line is drawn.

3. Results

Following a standard clinical approach, risk maps were evaluated by looking for the highest correlation threshold still containing activation, as explained in 2.5.2. Using this approach one value was collected for each combination of conditions. With the ROC curve analysis on the other hand, each voxel in each ROI is evaluated related to its status as TP or FP and its corresponding z-value (explained in 2.5.4.), leading to a much higher number of data points.

3.1. Comparing risk maps of cortex and brainstem

In the first step risk maps of cortex and brainstem were compared in order to get an overview of the data situation of brainstem in relation to cortex.

3.1.1. Cortex:

Table 2. Frequencies of highest correlation thresholds with activation in motor cortex.

Correlation Threshold		Frequency	Percent	Valid Percent
Valid	.25	1	,3	,5
	.30	5	1,7	2,3
	.35	10	3,5	4,6
	.40	9	3,1	4,2
	.45	6	2,1	2,8
	.50	19	6,6	8,8
	.55	20	6,9	9,3
	.60	25	8,7	11,6
	.65	31	10,8	14,4
	.70	24	8,3	11,1
	.75	41	14,2	**19,0**
	.80	25	8,7	11,6
	Total	216	75,0	100,0
Missing	.00	72	25,0	
Total		288	100,0	

In total, there are 288 data points, where 72 are missing values as BS mask excludes the cortex. Therefore, our cortex risk maps contain 216 valid values, comprising: 6 subjects, 4 sessions per subject, 3 preprocessing conditions per session and 3 reference functions per condition.

As shown in Table 2, in 216 cases (=100%) an activation in the motor cortex can be observed. Maximum correlation thresholds of observed activations range from 0.25 - 0.8. The most frequently occurring maximal correlation threshold with activation is 0.75 with 19%, followed by 0.65 with 14,4%.

3.1.2. Brainstem:

Table 3. Frequencies of highest correlation thresholds with activation in brainstem.

Correlation Threshold		Frequency	Percent	Valid Percent
Valid	.20	4	1,4	3,9
	.25	9	3,1	8,7
	.30	5	1,7	4,9
	.35	17	5,9	16,5
	.40	19	6,6	18,4
	.45	11	3,8	10,7
	.50	16	5,6	15,5
	.55	14	4,9	13,6
	.60	7	2,4	6,8
	.65	1	,3	1,0
	Total	103	35,8	100,0
Missing	.00	185	64,2	
Total		288	100,0	

Brainstem risk maps comprise 288 possible valid data points. They include: 6 subjects, 4 sessions per subject, 4 preprocessing conditions per session and 3 reference functions per condition.

Out of all possible data points, 103 (= 35,8%) show activation in trigeminal motor nuclei. Maximal correlation thresholds with BOLD signals range from 0.2 - 0.65. Most frequently the maximal correlation threshold with activation was 0.4 with 18%, followed by 0.35 with 16.5%. (see Table 3)

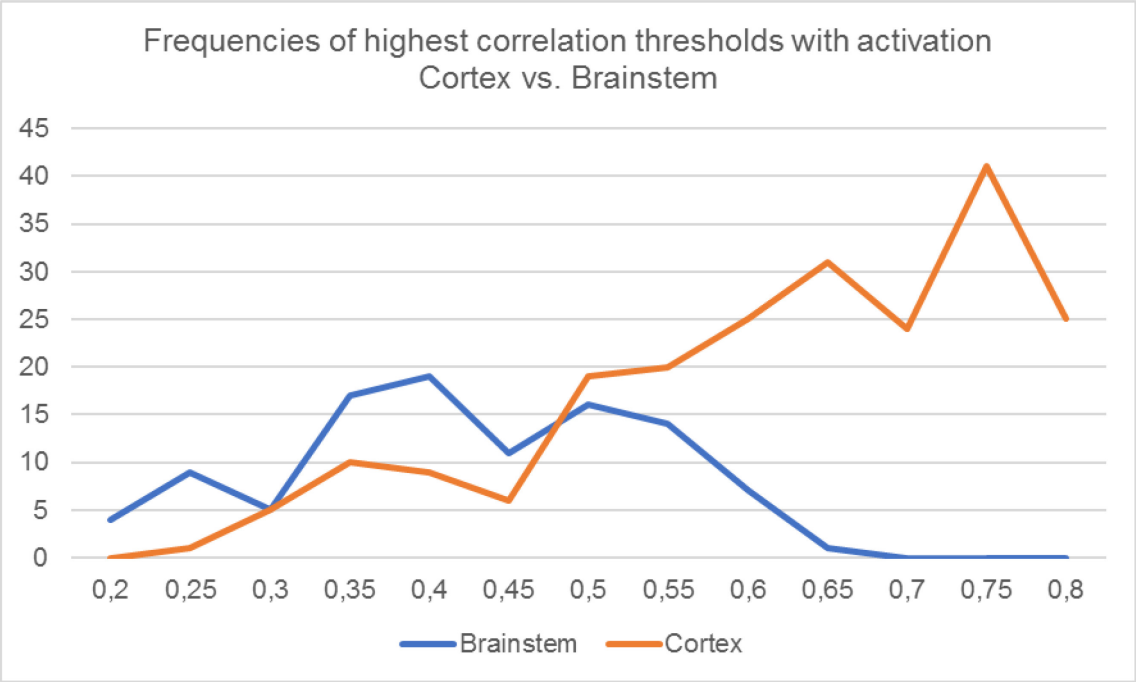

Figure 14. Frequencies of highest correlation thresholds with detectable BOLD signal in cortex vs. brainstem. The figure shows correlation thresholds on the x-axis and frequency of activation on the y-axis.

In Figure 14 cortex and brainstem curves, showing the frequencies of highest correlation thresholds with activation, are overlapped. Motor cortex activation can be detected more frequently at correlation thresholds between 0.5 and 0.8. Activations in trigeminal motor nuclei are less frequently detectable (36% of all cases) and in lower correlation thresholds than motor cortex. This indicates the relevant noise contamination of BOLD signals within the brainstem.

In the next steps, we were trying to assess if, despite of the lower correlation thresholds, reliable BOLD signals of brainstem scans can be gained or rather, which condition gives the most reliable BOLD signals, using risk maps.

3.2. Comparing influence of different preprocessing conditions on results of risk maps and ROC curves in brainstem scans

3.2.1. Risk map technique:

Figure 15 gives an overview of the risk map results of trigeminal motor nuclei activation. Data are separated for the 4 preprocessing conditions and contain all subjects, sessions and reference functions. In this overview “No Mask, No Smoothing” leads to activation in lowest correlation thresholds. “No Mask, Smoothing” seems to fit best for our data, as there were most cases of activation in higher correlation thresholds.

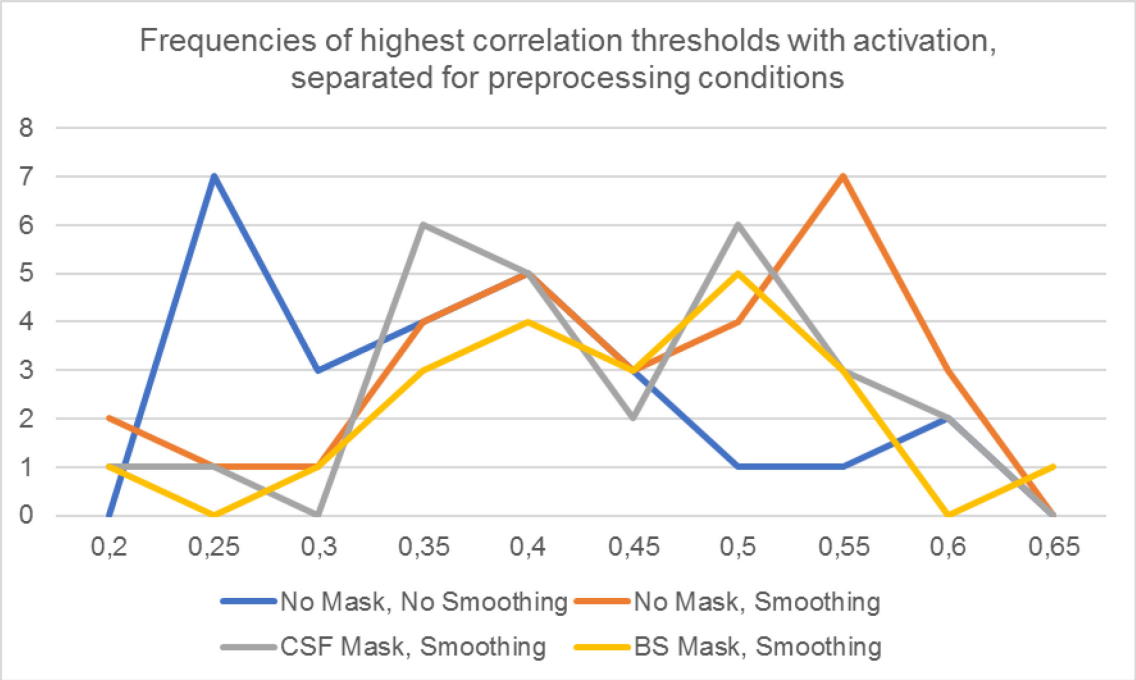

Figure 15. Frequencies of highest correlation thresholds with detectable BOLD signal in brainstem, separated for preprocessing conditions. The figure shows correlation thresholds on the x-axis and frequency of activation on the y-axis.

However, evaluating these results with a Chi-square test of independence yielded p>0,05 (see Table 4). Therefore, there is no significant association between maximal correlation thresholds with activation and the different preprocessing conditions.

Table 4. Results of Chi-square test of independence for correlation thresholds with activation and preprocessing conditions.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**34,465[a]**	**27**	**,153**
Likelihood Ratio	36,455	27	,106
Linear-by-Linear Association	4,212	1	,040
N of Valid Cases	103		

a. 39 cells (97,5%) have expected count less than 5. The minimum expected count is ,20.

Data situation can vary between subjects. With the difficulties of brainstem scans, individual differences may be even bigger. Therefore, the extent of individual differences was evaluated in the next step.

Table 5. Overview of risk map results of brainstem, separated for each subject.

	BS01	BS02	BS03	BS04	BS05	BS06
Valid Cases	5 (10,4%)	32 (66,7%)	19 (39,6%)	13 (27,1%)	19 (39,6%)	15 (31,3%)
Missing Cases	43 (89,6%)	16 (33,3%)	29 (60,4%)	35 (72,9%)	29 (60,4%)	33 (68,8%
Minimum	.20	.25	.45	.25	.40	.25
Maximum	.25	.45	.65	.40	.60	.55

Table 5 gives an overview of our brainstem risk map results separated for each subject. Shown are valid and missing cases as well as minimal and maximal value of highest correlation threshold with activation. Looking at the data of subject BS01, 90% are missing cases, meaning that only in 10% of all cases activation could be observed in trigeminal motor nuclei. Furthermore, maximal correlation thresholds of the activation detected in this subject are 0.2 and 0.25, which are the lowest possible. Subject BS02 on the other hand had detectable BOLD signals in 67% of the cases. Highest correlation thresholds with activation could be observed with subject BS03. This illustrates the considerable between-subject variability already in healthy subjects. Quite evidently, it may be expected that this variability considerably increases with patients.

Table 6. Results of Chi-square test of independence for correlation thresholds with activation and differrent subjects.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**199,963[a]**	**45**	**,000**
Likelihood Ratio	163,110	45	,000
N of Valid Cases	103		

a. 58 cells (96,7%) have expected count less than 5. The minimum expected count is ,05.

Additionally, a Chi-Square test for correlation thresholds with activation and different subjects was done, with an outcome of $p<0.001$ (see Table 6). Therefore, there is an association between maximal correlation threshold with activation and the different subjects. Maximal correlation thresholds with activation are significantly different, depending on the subject.

As maximal correlation thresholds with activation are different for each subject, we assessed the influence of the different preprocessing conditions separately for each subject.

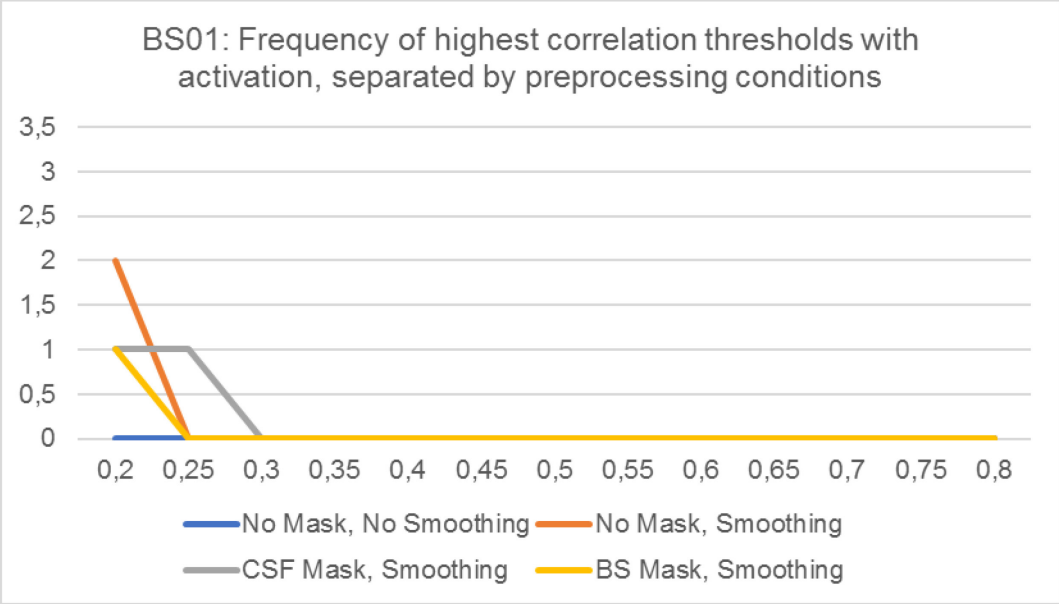

Figure 16. Frequency of highest correlation thresholds with activation, separated by preprocessing conditions of subject BS01. The figure shows correlation thresholds on the x-axis and frequency of activation on the y-axis.

As already shown in Table 5, not much activation could be extracted from the risk maps of subject BS01. “No Mask, No Smoothing” gave no activation at all. For the other 3 conditions, no trend can be seen (see Figure 16).

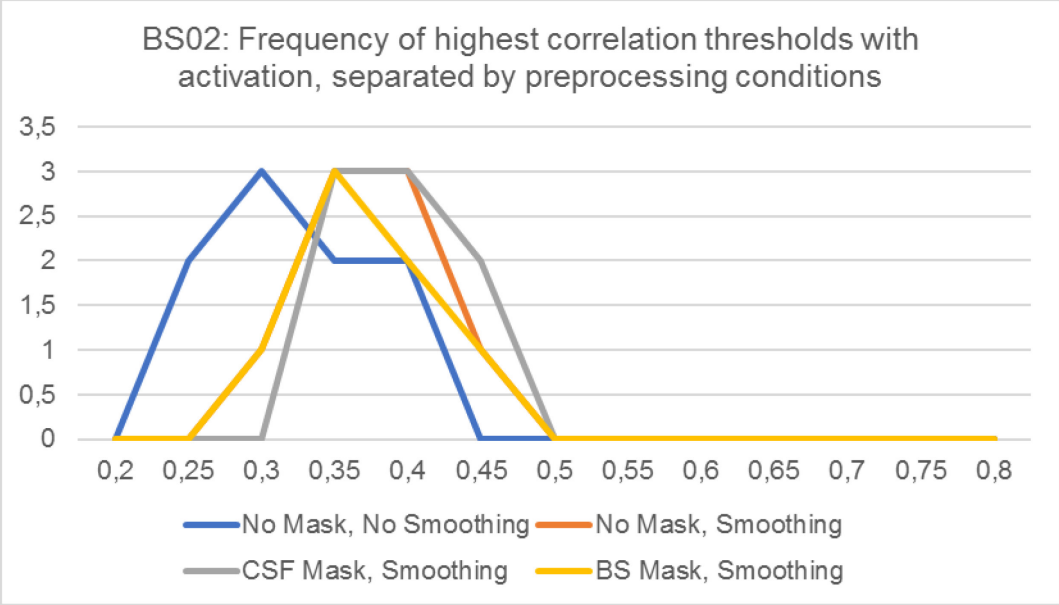

Figure 17. Frequency of highest correlation thresholds with activation, separated by preprocessing conditions of subject BS02. The figure shows correlation thresholds on the x-axis and frequency of activation on the y-axis.

Figure 17 shows that in case of subject BS02 “No Mask, No Smoothing” leads to BOLD signals in lower correlation thresholds than the other conditions. Regarding the other 3 conditions, “CSF Mask” yields slightly more activation.

In Figure 18 results of subject BS03 are shown. Maximal correlation thresholds with activation for all 4 conditions start at 0.45, which is higher than in the subjects before. However, “No Mask, No Smoothing” yields activation less frequently and on lowest correlation thresholds than the other conditions. The results of the other 3 conditions fluctuate, there is no clear trend.

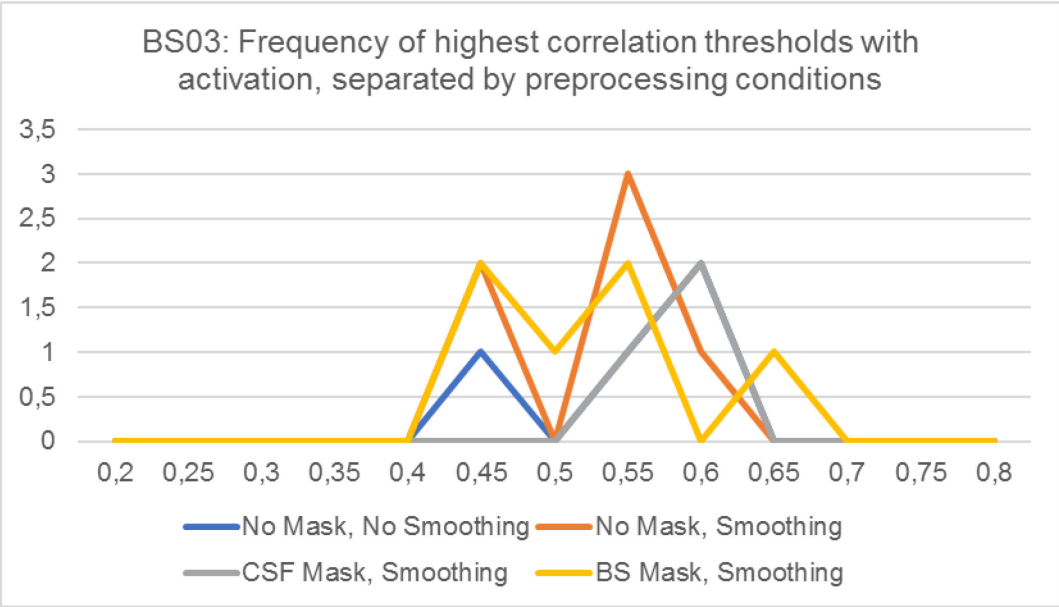

Figure 18. Frequency of highest correlation thresholds with activation, separated by preprocessing conditions of subject BS03. The figure shows correlation thresholds on the x-axis and frequency of activation on the y-axis.

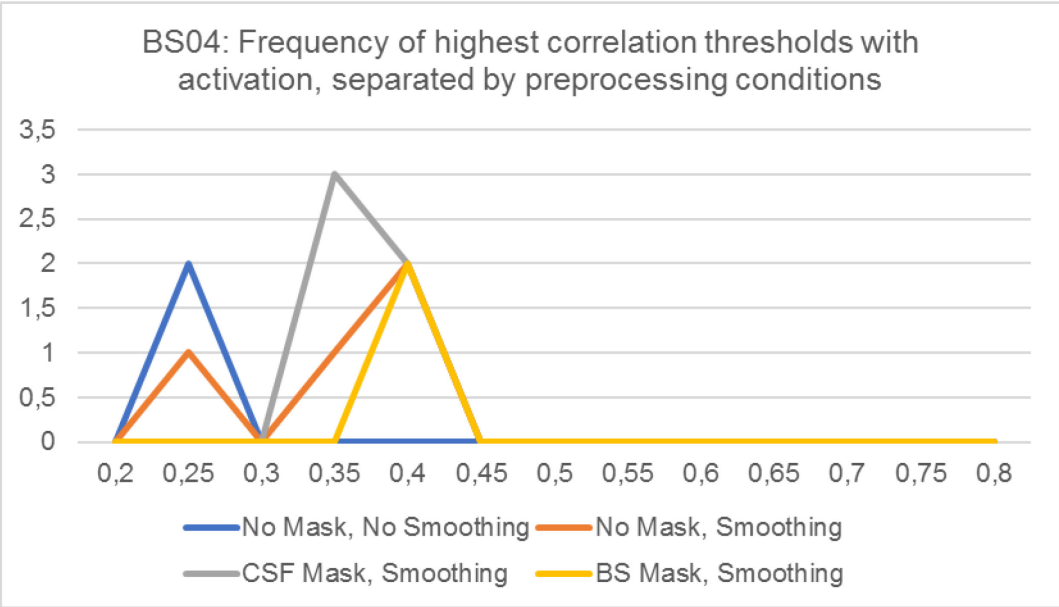

Figure 19. Frequency of highest correlation thresholds with activation, separated by preprocessing conditions of subject BS04. The figure shows correlation thresholds on the x-axis and frequency of activation on the y-axis.

Also in case of subject BS04 “No Mask, No Smoothing” leads to activation in lowest correlation thresholds. Regarding the other 3 conditions, no clear differences can be seen (see Figure 19).

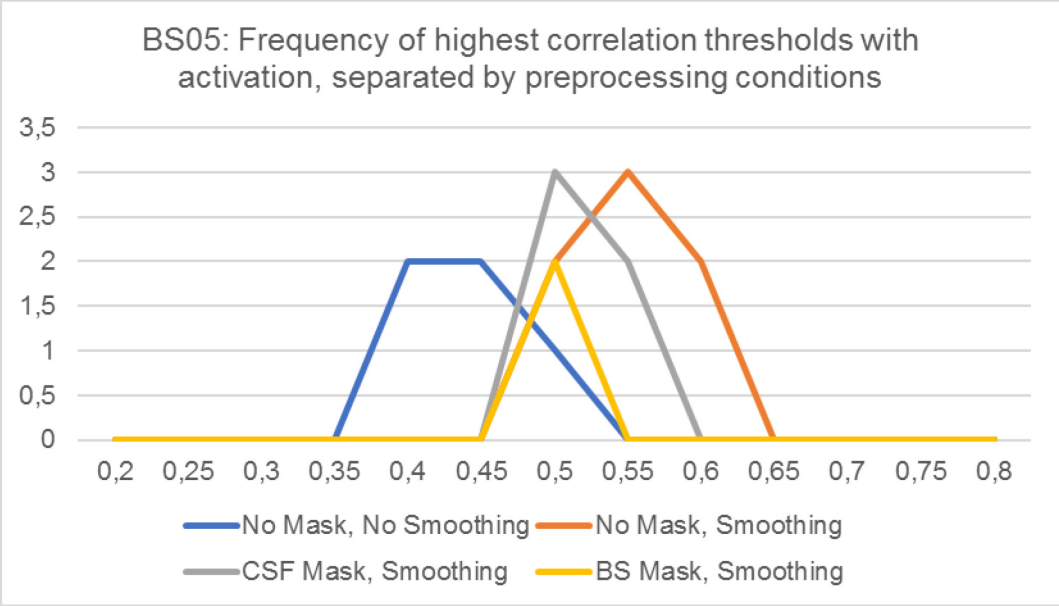

Figure 20. Frequency of highest correlation thresholds with activation, separated by preprocessing conditions of subject BS05. The figure shows correlation thresholds on the x-axis and frequency of activation on the y-axis.

Maximal correlation thresholds with activation in scans of subject BS05 are lowest with “No Mask, No Smoothing”. “No Mask, Smoothing” yields highest activation thresholds. (see Figure 20)

Risk maps of subject BS06 show lowest correlation thresholds with activation using condition “No Mask, No Smoothing”. The other conditions lead to similar results (see Figure 21).

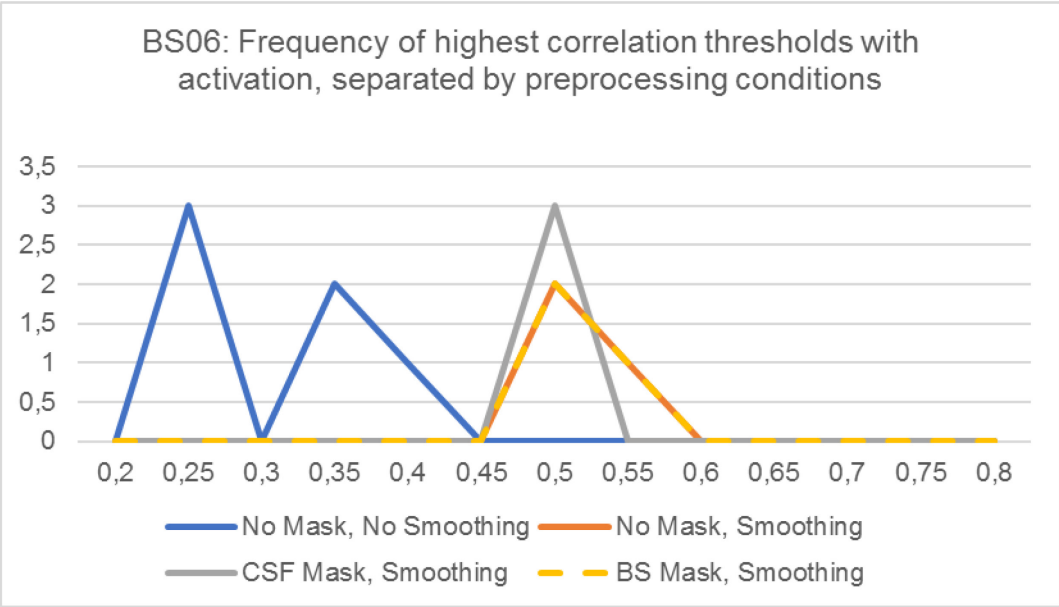

Figure 21. Frequency of highest correlation thresholds with activation, separated by preprocessing conditions of subject BS06. The figure shows correlation thresholds on the x-axis and frequency of activation on the y-axis.

Evaluating these results with a Chi-square test showed, that in none of the subjects a significant association between maximal correlation thresholds with activation and preprocessing conditions can be found (see Table 7).

Table 7. Results of Chi-square test of independence for maximal correlation thresholds with activation and preprocessing conditions for each subject.

Subject	Pearson Chi² Sign.
BS01	0.392 p > 0.05
BS02	0.495 p > 0.05
BS03	0.588 p > 0.05
BS04	0.070 p > 0.05
BS05	0.060 p > 0.05
BS06	0.144 p > 0.05

Taken together, condition “No Mask, No Smoothing” leads to activation in lowest correlation thresholds in all 6 subjects. However, the results using “No Mask, No Smoothing” are not significantly different from the other conditions, even though a trend can be seen. Further, it is

not clear which of the other 3 conditions was most beneficial to show activation of our images. This outcome is different to the group evaluation, where data of all subjects together were shown, and “No Mask, Smoothing” seemed to show activation best. This further illustrates that individual analyses as opposed to group analyses may lead to different inferences with fMRI data (Vandenbroucke et al., 2004).

3.2.2. ROC curve analysis:

In the next step, we did the same evaluations using ROC curves. As already mentioned, risk maps are evaluated visually by a person. ROC curves on the other hand are calculated by the computer, and were used to assess the quality of our data and discriminatory power of our tests.

Table 8. Results of ROC curve analysis, separately for preprocessing conditions.

		No Mask, No Smoothing	No Mask, Smoothing	CSF Mask, Smoothing	BS Mask, Smoothing
HRF	AUC	0.609	0.606	0.628	0.621
	Conf. Interv.	0.585-0.634	0.583-0.629	0.605-0.651	0.598-0.644
Shift +0	AUC	0.594	0.571	0.595	0.588
	Conf. Interv.	0.569-0.618	0.548-0.595	0.572-0.618	0.565-0.612
Shift +1	AUC	0.609	0.617	0.64	0.635
	Conf. Interv..	0.585-0.634	0.593-0.640	0.617-0.662	0.612-0.658

Looking at the influence of the different preprocessing conditions on our data over all subjects and sessions, “CSF Mask, Smoothing” yields highest AUCs, followed by “BS Mask, Smoothing”. Both were higher than the AUCs of the other conditions (with one exception, see Table 8). As we now know about the high individual differences, data were separated for each subject in the next step.

Table 9. Results of ROC curve analysis for subject BS01, separately for preprocessing conditions.

BS01		No Mask, No Smoothing	No Mask, Smoothing	CSF Mask, Smoothing	BS Mask, Smoothing
HRF	AUC	0.491	0.451	0.479	0.482
	Conf. Interv.	0.423-0.559	0.388-0.514	0.416-0.542	0.419-0.545
Shift +0	AUC	0.491	0.457	0.481	0.490
	Conf. Interv.	0.422-0.559	0.394-0.519	0.419-0.542	0.428-0.552
Shift +1	AUC	0.490	0.450	0.493	0.497
	Conf. Interv.	0.422-0.558	0.387-0.512	0.431-0.555	0.435-0.559

Images of subject BS01 have low data quality according to ROC curve analysis, which could also be seen in the risk map analysis. Values in Table 9 are grey, as they are not significantly different from the 45° reference line. Therefore, our tests are not specific and sensitive enough to distinguish between TP and FP values. An example of these ROC curves without discriminatory power can be seen in Figure 22.

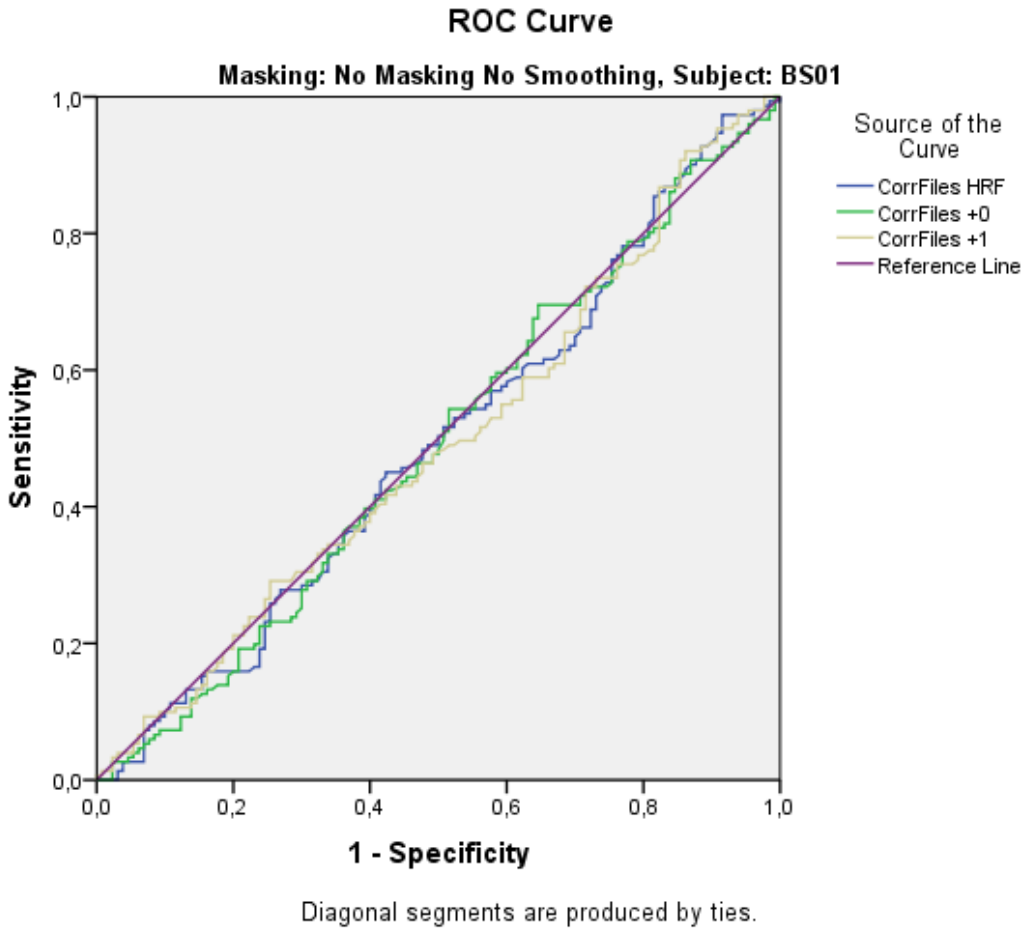

Figure 22. Example of ROC curves without discriminatory power. This figure shows ROC curves of subject BS01 for condition “No Mask, No Smoothing”. The three lines represent the three reference functions “HRF”, “Shift+0” and “Shift +1”. Additionally a 45° reference line is drawn. All three lines lie very close to the 45° reference line, therefore our test is not specific and sensitive enough to distinguish between TP and FP values.

Table 10. Results of ROC curve analysis for subject BS02, separately for preprocessing conditions.

BS02		No Mask, No Smoothing	No Mask, Smoothing	CSF Mask, Smoothing	BS Mask, Smoothing
HRF	AUC	0.677	0.786	0.821	0.812
	Conf. Interv.	0.624-0.729	0.747-0.825	0.785-0.857	0.776-0.849
Shift +0	AUC	0.680	0.760	0.810	0.795
	Conf. Interv.	0.627-0.733	0.719-0.801	0.772-0.847	0.757-0.833
Shift +1	AUC	0.657	0.783	0.789	0.811
	Conf. Interv.	0.604-0.711	0.744-0.822	0.750-0.828	0.774-0.847

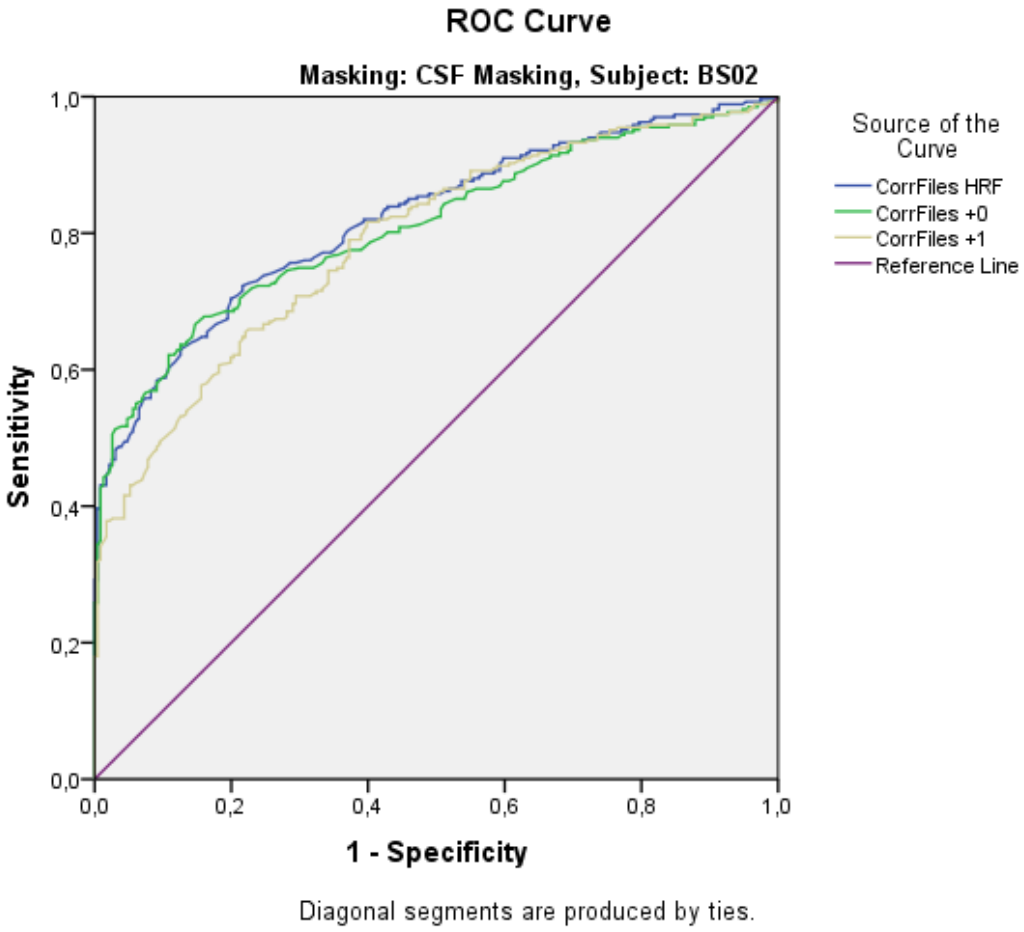

Figure 23. Example of ROC curves with discriminatory power. This figure shows ROC curves of subject BS02 for condition "CSF Mask, Smoothing". The three lines represent the three reference functions "HRF", "Shift+0" and "Shift +1". Additionally a 45° reference line is drawn. All three lines have high AUCs, they lie above the 45° reference line, therefore our test is specific and sensitive enough to distinguish between TP and FP values.

AUCs of subject BS02 show that „CSF Mask, Smoothing“, and „BS Mask, Smoothing“ are the best conditions to evaluate these data. “No Mask, No Smoothing” yields the lowest AUCs (see Table 10). In case of BS02 we seem to have good data quality, as AUCs go up to above 0.8, an example of these curves is shown in Figure 23.

Table 11. Results of ROC curve analysis for subject BS03, separately for preprocessing conditions.

BS03		No Mask, No Smoothing	No Mask, Smoothing	CSF Mask, Smoothing	BS Mask, Smoothing
HRF	AUC	0.710	0.763	0.742	0.746
	Conf. Interv.	0.638-0.783	0.662-0.863	0.647-0.838	0.645-0.847
Shift +0	AUC	0.698	0.689	0.694	0.676
	Conf. Interv.	0.624-0.772	0.578-0.800	0.591-0.796	0.566-0.785
Shift +1	AUC	0.707	0.721	0.719	0.732
	Conf. Interv.	0.634-0.781	0.613-0.828	0.619-0.819	0.629-0.835

Table 11 shows that AUCs of subject BS03 fluctuate over the different conditions. In 2 of 3 reference functions “No Mask, No Smoothing” has lowest AUCs.

Table 12. Results of ROC curve analysis for subject BS04, separately for preprocessing conditions.

BS04		No Mask, No Smoothing	No Mask, Smoothing	CSF Mask, Smoothing	BS Mask, Smoothing
HRF	AUC	0.726	0.786	0.850	0.808
	Conf. Interv.	0.673-0.778	0.745-0.827	0.817-0.883	0.769-0.847
Shift +0	AUC	0.665	0.691	0.766	0.718
	Conf. Interv.	0.610-0.721	0.643-0.739	0.724-0.807	0.672-0.764
Shift +1	AUC	0.746	0.831	0.887	0.844
	Conf. Interv.	0.695-0.796	0.795-0.868	0.859-0.916	0.809-0.879

In case of subject BS04, condition “No Mask, No Smoothing” yields lowest AUCs. The highest sensitivity and specificity could be gained with “CSF Mask, Smoothing”. Also, data quality seems to be high in these scans (see Table 12).

Table 13. Results of ROC curve analysis for subject BS05, separately for preprocessing conditions.

BS05		No Mask, No Smoothing	No Mask, Smoothing	CSF Mask, Smoothing	BS Mask, Smoothing
HRF	AUC	0.619	0.642	0.578	0.639
	Conf. Interv.	0.569-0.668	0.569-0.688	0.530-0.626	0.593-0.684
Shift +0	AUC	0.616	0.616	0.542	0.610
	Conf. Interv.	0.566-0.666	0.569-0.663	0.493-0.591	0.563-0.656
Shift +1	AUC	0.608	0.632	0.593	0.640
	Conf. Interv.	0.558-0.658	0.586-0.678	0.545-0.640	0.595-0.686

In case of subject BS05, “CSF Mask, Smoothing” leads to lowest AUCs. Values of the other 3 conditions fluctuate (see Table 13).

Table 14. Results of ROC curve analysis for subject BS06, separately for preprocessing conditions.

BS06		No Mask, No Smoothing	No Mask, Smoothing	CSF Mask, Smoothing	BS Mask, Smoothing
HRF	AUC	0.628	0.574	0.625	0.617
	Conf. Interv.	0.569-0.687	0.512-0.635	0.567-0.682	0.560-0.674
Shift +0	AUC	0.609	0.542	0.612	0.599
	Conf. Interv.	0.549-0.669	0.480-0.604	0.553-0.670	0.541-0.656
Shift +1	AUC	0.646	0.629	0.682	0.673
	Conf. Interv.	0.588-0.704	0.569-0.689	0.626-0.737	0.617-0.728

Condition “No Mask, Smoothing” yields lowest AUCs in images of subject BS06. “CSF Mask, Smoothing” seems to show a better performance than the other test conditions (see Table 14).

Summarized, performances of the different test conditions separated for each subject vary and don’t show a clear trend of which condition performs worst or best. As we have seen high individual differences in our data with risk maps and ROC curves, we wanted to know if data quality also varies between sessions.

Table 15. Results of ROC cuve analysis of each subject and session.

		Sess. 1	Sess. 2	Sess. 3	Sess. 4
BS01	HRF	0.232	0.866	0.484	0.715
	Shift +0	0.209	0.753	0.570	0.604
	Shift +1	0.248	0.692	0.541	0.382
BS02	HRF	0.710	0.779	0.812	0.768
	Shift +0	0.695	0.752	0.771	0.746
	Shift +1	0.701	0.772	0.813	0.741
BS03	HRF	0.658	0.824	0.683	0.570
	Shift +0	0.723	0.787	0.614	0.512
	Shift +1	0.693	0.781	0.687	0.463
BS04	HRF	0.745	0.771	0.843	0.845
	Shift +0	0.728	0.658	0.790	0.803
	Shift +1	0.793	0.845	0.890	0.803
BS05	HRF	0.727	0.472	0.663	0.503
	Shift +0	0.685	0.451	0.675	0.462
	Shift +1	0.721	0.459	0.667	0.516
BS06	HRF	0.916	0.612	0.572	0.405
	Shift +0	0.888	0.585	0.525	0.409
	Shift +1	0.903	0.577	0.599	0.622

Table 15 shows not only individual, but also inter-individual differences. Images of subjects BS01, BS05 and BS06 seem to have poor data quality in 2-3 sessions. Scans of BS02 and BS04 on the other hand resulted in good data quality in all 4 sessions. However, data quality of our scans varies between subjects and sessions. Therefore, ROC curves for the different conditions were made for each session of each subject in the next step. It would not have made sense to apply this approach to the risk map results, as there are much fewer data points.

Table 16. Results of ROC curve analysis for each session of subject BS01, separately for preprocessing conditions.

BS01			No Mask, No Smoothing	No Mask, Smoothing	CSF Mask, Smoothing	BS Mask, Smoothing
Session 1	HRF	AUC	0.318	0.185	0.212	0.179
		Conf. Interv.	0.201-0.436	0.106-0.264	0.128-0.296	0.104-0.254
	Shift +0	AUC	0.304	0.181	0.190	0.180
		Conf. Interv.	0.190-0.418	0.104-0.257	0.111-0.269	0.104-0.256
	Shift +1	AUC	0.332	0.195	0.236	0.204
		Conf. Interv.	0.211-0.453	0.108-0.282	0.143-0.330	0.120-0.288
Session 2	HRF	AUC	0.705	0.916	0.894	0.918
		Conf. Interv.	0.554-0.855	0.843-0.990	0.815-0.972	0.850-0.987
	Shift +0	AUC	0.630	0.775	0.782	0.823
		Conf. Interv.	0.479-0.782	0.657-0.893	0.674-0.890	0.722-0.924
	Shift +1	AUC	0.615	0.682	0.727	0.783
		Conf. Interv.	0.467-0.762	0.551-0.812	0.607-0.848	0.668-0.898
Session 3	HRF	AUC	0.490	0.471	0.497	0.509
		Conf. Interv.	0.372-0.608	0.375-0.566	0.401-0.593	0.412-0.605
	Shift +0	AUC	0.551	0.569	0.601	0.625
		Conf. Interv.	0.434-0.669	0.474-0.664	0.508-0.695	0.532-0.719
	Shift +1	AUC	0.553	0.519	0.557	0.578
		Conf. Interv.	0.434-0.672	0.424-0.615	0.462-0.652	0.483-0.673
Session 4	HRF	AUC	0.495	0.958	0.939	0.970
		Conf. Interv.	0.307-0.683	0.839-1.000	0.801-1.000	0.882-1.000
	Shift +0	AUC	0.495	0.750	0.727	0.780
		Conf. Interv.	0.308-0.682	0.432-1.000	0.413-1.000	0.477-1.000
	Shift +1	AUC	0.471	0.375	0.318	0.333
		Conf. Interv.	0.289-0.653	0.000-0.821	0.000-0.719	0.000-0.754

Shown in Table 16 are the results of ROC curve analysis of subject BS01 for each session, separated for the 4 preprocessing conditions. Values in grey are not significantly different from the 45° reference line, therefore it is not possible to distinguish between TP and FP in these cases. Data quality of images of subject BS01 is poor in many cases (grey values). Further, “No Mask, No Smoothing” is not suitable to show these data reliably. Regarding the other 3 conditions “BS Mask, Smoothing” yields highest AUC values.

Table 17. Results of ROC curve analysis for each session of subject BS02, separately for preprocessing conditions.

BS02			No Mask, No Smoothing	No Mask, Smoothing	CSF Mask, Smoothing	BS Mask, Smoothing
Session 1	HRF	AUC	0.584	0.707	0.740	0.766
		Conf. Interv.	0.584-0.799	0.608-0.806	0.644-0.836	0.676-0.856
	Shift +0	AUC	0.585	0.691	0.720	0.748
		Conf. Interv.	0.585-0.809	0.590-0.791	0.620-0.819	0.655-0.840
	Shift +1	AUC	0.557	0.708	0.693	0.754
		Conf. Interv.	0.557-0.782	0.609-0.806	0.592-0.794	0.662-0.846
Session 2	HRF	AUC	0.656	0.810	0.849	0.780
		Conf. Interv.	0.550-0.763	0.741-0.879	0.787-0.911	0.780-0.906
	Shift +0	AUC	0.698	0.765	0.838	0.755
		Conf. Interv.	0.595-0.801	0.688-0.842	0.773-0.902	0.755-0.890
	Shift +1	AUC	0.625	0.785	0.840	0.751
		Conf. Interv.	0.515-0.734	0.711-0.858	0.777-0.904	0.751-0.886
Session 3	HRF	AUC	0.746	0.827	0.838	0.841
		Conf. Interv.	0.642-0.850	0.757-0.896	0.768-0.908	0.775-0.907
	Shift +0	AUC	0.756	0.787	0.796	0.805
		Conf. Interv.	0.655-0.857	0.710-0.864	0.718-0.875	0.732-0.879
	Shift +1	AUC	0.753	0.839	0.813	0.850
		Conf. Interv.	0.652-0.853	0.772-0.905	0.738-0.887	0.786-0.914
Session 4	HRF	AUC	0.571	0.837	0.899	0.846
		Conf. Interv.	0.461-0.680	0.769-0.905	0.847-0.950	0.780-0.912
	Shift +0	AUC	0.554	0.817	0.895	0.825
		Conf. Interv.	0.444-0.664	0.743-0.890	0.841-0.949	0.754-0.896
	Shift +1	AUC	0.554	0.813	0.802	0.839
		Conf. Interv.	0.440-0.667	0.742-0.884	0.729-0.874	0.773-0.904

AUCs of each session of BS02, separated for preprocessing conditions, are shown in Table 17. In this case we see can good data quality in all sessions. Condition “No Mask, No Smoothing” yields lowest AUCs. Further, for each session either “CSF Mask, Smoothing”, or “BS Mask, Smoothing” leads to highest AUCs.

Table 18. Results of ROC curve analysis for each session of subject BS03, separately for preprocessing conditions.

BS03			No Mask, No Smoothing	No Mask, Smoothing	CSF Mask, Smoothing	BS Mask, Smoothing
Session 1	HRF	AUC	0.706			
		Conf. Interv.	0.504-0.909			
	Shift +0	AUC	0.788			
		Conf. Interv.	0.611-0.964			
	Shift +1	AUC	0.684			
		Conf. Interv.	0.484-0.884			
Session 2	HRF	AUC	0.762	0.982	0.893	0.928
		Conf. Interv.	0.637-0.886	0.947-1.000	0.799-0.987	0.854-1.000
	Shift +0	AUC	0.756	0.923	0.861	0.840
		Conf. Interv.	0.629-0.882	0.843-1.000	0.764-0.957	0.740-0.940
	Shift +1	AUC	0.739	0.836	0.777	0.824
		Conf. Interv.	0.607-0.871	0.728-0.945	0.667-0.886	0.715-0.933
Session 3	HRF	AUC	0.668	0.800	0.667	0.900
		Conf. Interv.	0.506-0.830	0.449-1.000	0.222-1.000	0.644-1.000
	Shift +0	AUC	0.650	0.300	0.400	0.700
		Conf. Interv.	0.488-0.811	0.000-0.697	0.000-0.850	0.303-1.000
	Shift +1	AUC	0.663	0.900	0.733	1.000
		Conf. Interv.	0.500-0.826	0.644-1.000	0.81-1.000	01.000-1.000
Session 4	HRF	AUC	0.600			
		Conf. Interv.	0.437-0.764			
	Shift +0	AUC	0.556			
		Conf. Interv.	0.387-0.725			
	Shift +1	AUC	0.598			
		Conf. Interv.	0.433-0.764			

Regarding subject BS03, only session 2 gave results where sensitivity and specificity were high enough to distinguish TP from FP. “No Mask, No Smoothing” resulted in lowest AUCs. “No Mask, Smoothing” seems to show data best in this case (see Table 18).

Table 19. Results of ROC curve analysis for each session of subject BS04, separately for preprocessing conditions.

BS04			No Mask, No Smoothing	No Mask, Smoothing	CSF Mask, Smoothing	BS Mask, Smoothing
Session 1	HRF	AUC	0.586	0.785	0.824	0.789
		Conf. Interv.	0.454-0.718	0.693-0.876	0.742-0.905	0.698-0.880
	Shift +0	AUC	0.526	0.755	0.810	0.757
		Conf. Interv.	0.390-0.663	0.660-0.851	0.727-0.892	0.663-0.852
	Shift +1	AUC	0.662	0.820	0.883	0.818
		Conf. Interv.	0.490-0.753	0.736-0.904	0.818-0.947	0.734-0.901
Session 2	HRF	AUC	0.720	0.840	0.891	0.872
		Conf. Interv.	0.621-0.818	0.774-0.906	0.836-0.947	0.812-0.932
	Shift +0	AUC	0.609	0.688	0.778	0.744
		Conf. Interv.	0.501-0.716	0.598-0.778	0.699-0.857	0.661-0.827
	Shift +1	AUC	0.752	0.890	0.926	0.915
		Conf. Interv.	0.659-0.845	0.838-0.942	0.882-0.970	0.870-0.960
Session 3	HRF	AUC	0.803	0.852	0.911	0.887
		Conf. Interv.	0.716-0.890	0.789-0.915	0.862-0.960	0.832-0.941
	Shift +0	AUC	0.793	0.780	0.873	0.836
		Conf. Interv.	0.702-0.883	0.700-0.861	0.809-0.936	0.767-0.905
	Shift +1	AUC	0.845	0.894	0.936	0.917
		Conf. Interv.	0.769-0.920	0.842-0.945	0.896-0.975	0.872-0.962
Session 4	HRF	AUC	0.780	0.833	0.927	0.861
		Conf. Interv.	0.678-0.886	0.751-0.915	0.880-0.975	0.788-0.934
	Shift +0	AUC	0.748	0.790	0.889	0.803
		Conf. Interv.	0.638-0.859	0.697-0.883	0.825-0.953	0.715-0.891
	Shift +1	AUC	0.741	0.794	0.868	0.801
		Conf. Interv.	0.628-0.855	0.702-0.886	0.800-0.936	0.713-0.888

In case of subject BS04 data quality seems to be good. “No Mask, No Smoothing” is the condition with lowest AUCs, whereas “CSF Mask, Smoothing” yields highest AUCs (see Table 19).

Table 20. Results of ROC curve analysis for each session of subject BS05, separately for preprocessing conditions.

BS05			No Mask, No Smoothing	No Mask, Smoothing	CSF Mask, Smoothing	BS Mask, Smoothing
Session 1	HRF	AUC	0.689	0.841	0.792	0.874
		Conf. Interv.	0.593-0.784	0.773-0.908	0.718-0.867	0.817-0.931
	Shift +0	AUC	0.652	0.796	0.718	0.823
		Conf. Interv.	0.554-0.751	0.720-0.871	0.633-0.804	0.755-0.891
	Shift +1	AUC	0.673	0.832	0.793	0.874
		Conf. Interv.	0.577-0.769	0.762-0.903	0.720-0.865	0.815-0.932
Session 2	HRF	AUC	0.561	0.462	0.372	0.405
		Conf. Interv.	0.456-0.667	0.3364-0.559	0.277-0.467	0.311-0.500
	Shift +0	AUC	0.546	0.432	0.334	0.387
		Conf. Interv.	0.440-0.651	0.335-0.528	0.241-0.426	0.293-0.482
	Shift +1	AUC	0.539	0.425	0.369	0.384
		Conf. Interv.	0.433-0.645	0.326-0.524	0.271-0.467	0.287-0.482
Session 3	HRF	AUC	0.696	0.726	0.667	0.737
		Conf. Interv.	0.602-0.789	0.640-0.812	0.570-0.764	0.651-0.823
	Shift +0	AUC	0.733	0.743	0.689	0.765
		Conf. Interv.	0.642-0.823	0.659-0.827	0.591-0.787	0.682-0.848
	Shift +1	AUC	0.677	0.713	0.689	0.733
		Conf. Interv.	0.580-0.774	0.623-0.804	0.592-0.786	0.643-0.823
Session 4	HRF	AUC	0.531	0.530	0.459	0.491
		Conf. Interv.	0.431-0.631	0.432-0.628	0.360-0.558	0.393-0.588
	Shift +0	AUC	0.519	0.471	0.396	0.416
		Conf. Interv.	0.418-0.619	0.374-0.569	0.298-0.494	0.320-0.512
	Shift +1	AUC	0.538	0.532	0.494	0.505
		Conf. Interv.	0.438-0.638	0.432-0.632	0.393-0.594	0.406-0.604

Table 20 shows results of subject BS05. In session 2 and 4 none of the conditions yields AUCs with enough discriminatory power to distinguish between TP and FP. In the other 2 sessions “BS Mask, Smoothing” is the condition with highest AUCs. “No Mask, No Smoothing” again leads to lowest AUCs.

Table 21. Results of ROC curve analysis for each session of subject BS06, separately for preprocessing conditions.

BS06			No Mask, No Smoothing	No Mask, Smoothing	CSF Mask, Smoothing	BS Mask, Smoothing
Session 1	HRF	AUC	0.841	0.973	0.983	0.984
		Conf. Interv.	0.749-0.933	0.931-1.000	0.950-1.000	0.953-1.000
	Shift +0	AUC	0.805	0.946	0.970	0.984
		Conf. Interv.	0.700-0.909	0.885-1.000	0.928-1.000	0.953-1.000
	Shift +1	AUC	0.853	0.982	0.979	0.984
		Conf. Interv.	0.766-0.940	0.947-1.000	0.944-1.000	0.953-1.000
Session 2	HRF	AUC	0.602	0.602	0.641	0.618
		Conf. Interv.	0.472-0.733	0.454-0.750	0.505-0.778	0.483-0.752
	Shift +0	AUC	0.577	0.571	0.609	0.589
		Conf. Interv.	0.445-0.709	0.419-0.723	0.464-0.754	0.449-0.730
	Shift +1	AUC	0.579	0.551	0.609	0.575
		Conf. Interv.	0.447-0.711	0.400-0.703	0.471-0.747	0.437-0.712
Session 3	HRF	AUC	0.535	0.549	0.637	0.590
		Conf. Interv.	0.424-0.647	0.439-0.660	0.533-0.741	0.484-0.697
	Shift +0	AUC	0.497	0.489	0.590	0.535
		Conf. Interv.	0.385-0.609	0.379-0.600	0.484-0.696	0.427-0.642
	Shift +1	AUC	0.525	0.576	0.674	0.614
		Conf. Interv.	0.412-0.637	0.465-0.686	0.572-0.776	0.508-0.720
Session 4	HRF	AUC	0.531	0.360	0.409	0.355
		Conf. Interv.	0.404-0.657	0.247-0.472	0.302-0.517	0.253-0.458
	Shift +0	AUC	0.538	0.351	0.417	0.362
		Conf. Interv.	0.412-0.664	0.243-0.459	0.312-0.522	0.262-0.462
	Shift +1	AUC	0.634	0.596	0.666	0.601
		Conf. Interv.	0.515-0.753	0.481-0.710	0.564-0.768	0.495-0.728

In Table 21 data of subject BS06 are shown. Data quality of images of Sessions 2, 3 and 4 is poor. Regarding session 1, “No Mask, No Smoothing” results in lowest AUCs, “BS Mask, Smoothing” in highest AUCs.

Looking at the different preprocessing conditions of each session of each subject, “No Mask, No Smoothing” leads to lowest AUCs. Often these AUCs are around 0.5, which means that it is not possible to distinguish between true positive and false positive BOLD signals. Concerning the other conditions, “CSF Mask, Smoothing” and “BS Mask, Smoothing” mostly yield highest AUCs, whereby with “BS Mask, Smoothing” this was more often the case.

3.3. Comparing influence of different reference functions on results of risk maps and ROC curves of brainstem scans

In the next part we wanted to assess, which reference function shows activations in our brainstem scans best. Therefore, we separated the results for the different reference functions, using risk maps as well as ROC curves.

3.3.1. Risk map technique:

Figure 24 shows that the reference function “Shift +1” seems most suitable to show activation of our brainstem risk maps, as it yields the highest frequency of BOLD signals in high correlation thresholds.

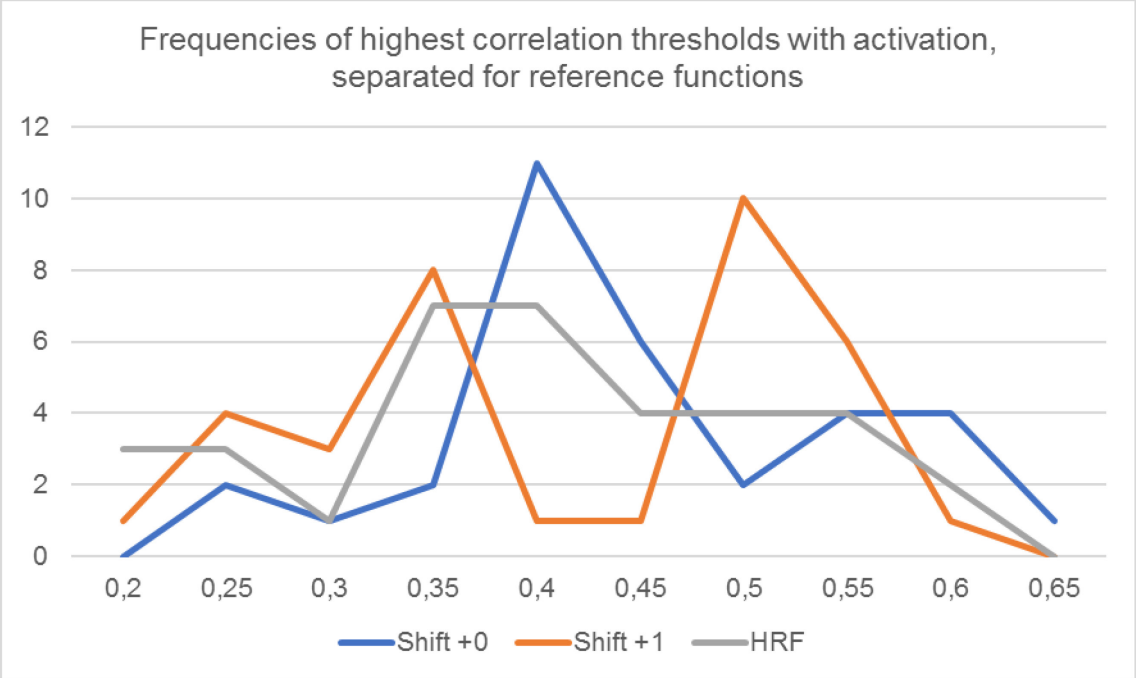

Figure 24. Frequencies of highest correlation thresholds with detectable BOLD signal in brainstem, lines separated for reference functions. The figure shows correlation thresholds on the x-axis and frequency of activation on the y-axis.

A Chi² test was performed, to check for associations between maximal correlation thresholds with activation and reference functions. The test resulted in $p < 0,05$. Therefore, there is a significant association between the two variables. Maximal correlation thresholds with activation are different, depending on the reference function used. (see Table 22)

Table 22. Results of Chi-square test of independence for correlation thresholds with activation and reference functions.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**32,080[a]**	**18**	**,022**
Likelihood Ratio	35,642	18	,008
Linear-by-Linear Association	3,075	1	,079
N of Valid Cases	103		

a. 21 cells (70,0%) have expected count less than 5. The minimum expected count is ,32.

In the next step, we assessed the influence of the different reference functions on our results, separated for preprocessing conditions, to see if separated for the preprocessing conditions, “Shift +1” also yields best results.

In Figure 25 results of our risk map evaluations using HRF function are shown. Results are separated for the 4 different preprocessing conditions, shown as the 4 different lines. The figure shows the frequency of the maximal threshold with activation in riskmaps. Using HRF as reference function, activation is present 0 – 2 times over the different correlation thresholds, except of threshold 0.25, where activation was counted 3 times.

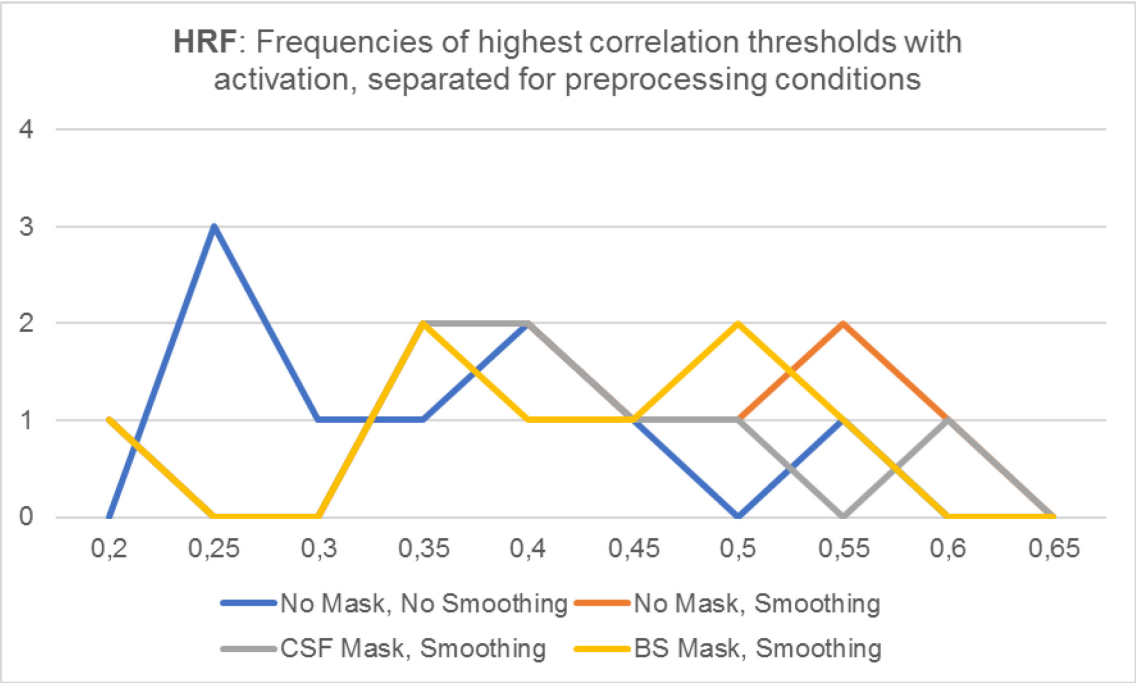

Figure 25. Frequencies of highest correlation thresholds with detectable BOLD signal in brainstem using "HRF". The figure shows correlation thresholds on the x-axis and frequency of activation on the y-axis. The lines show the four preprocessing conditions.

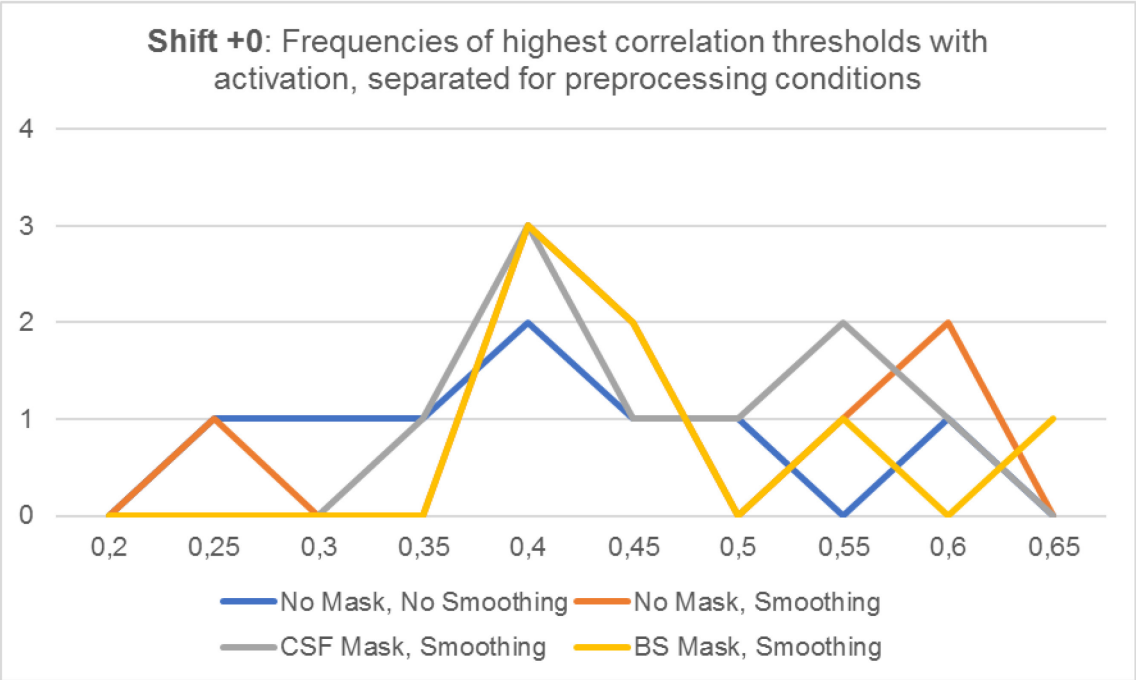

Figure 26. Frequencies of highest correlation thresholds with detectable BOLD signal in brainstem using "Shift +0". The figure shows correlation thresholds on the x-axis and frequency of activation on the y-axis. The lines show the four preprocessing conditions.

When reference function “Shift +0” is used on our risk map results, activation in the middle range (0,4 – 0,45) was counted more often (see Figure 26).

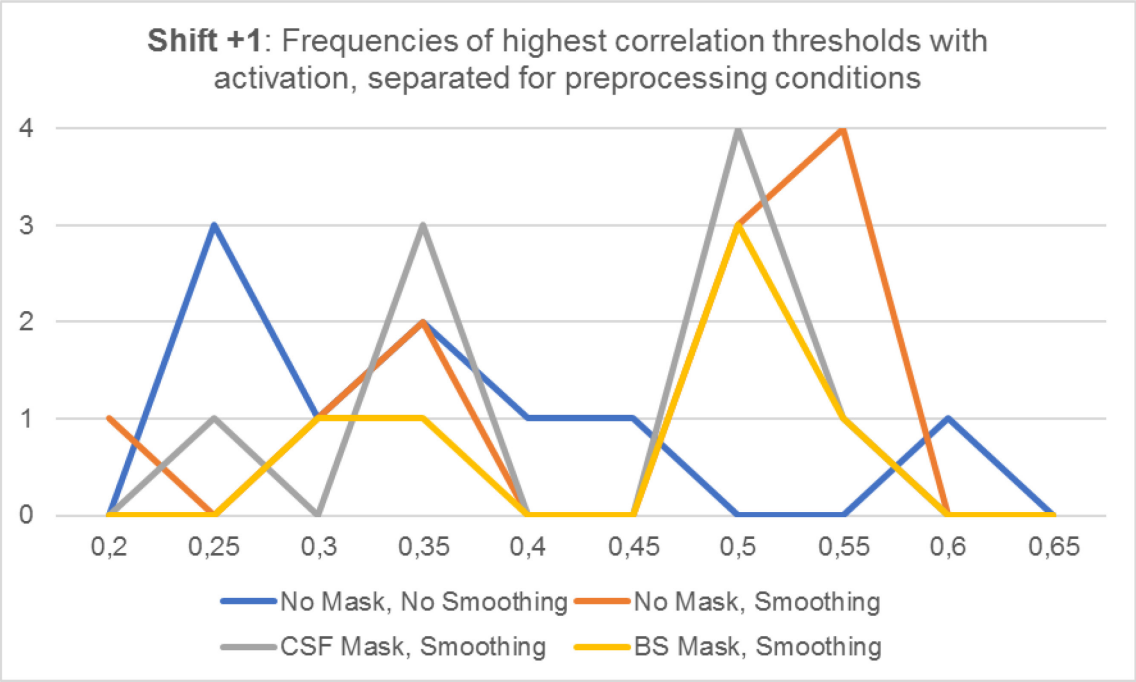

Figure 27 Frequencies of highest correlation thresholds with detectable BOLD signal in brainstem using “Shift +1”. The figure shows correlation thresholds on the x-axis and frequency of activation on the y-axis. The lines show the four preprocessing conditions.

Figure 27 shows risk map results, when Shift +1 is used as reference function. Activation is present more often in lower and in higher correlation thresholds. In thresholds 0,5 – 0,55 activations can be counted up to 4 times for 3 preprocessing conditions.

Comparing the three figures Figure 25, Figure 26 and Figure 27 showing performances of reference functions on our risk map data separated by preprocessing conditions, “Shift +1” yields activations more often at lower, but also, and more importantly, at higher correlation thresholds and therefore seems to fit best for our data.

3.3.2. ROC curve analysis:

The same comparison was done for the data of the ROC curve analysis. First, the influence of the different reference functions was examined over all data, then separated for the 4 preprocessing conditions.

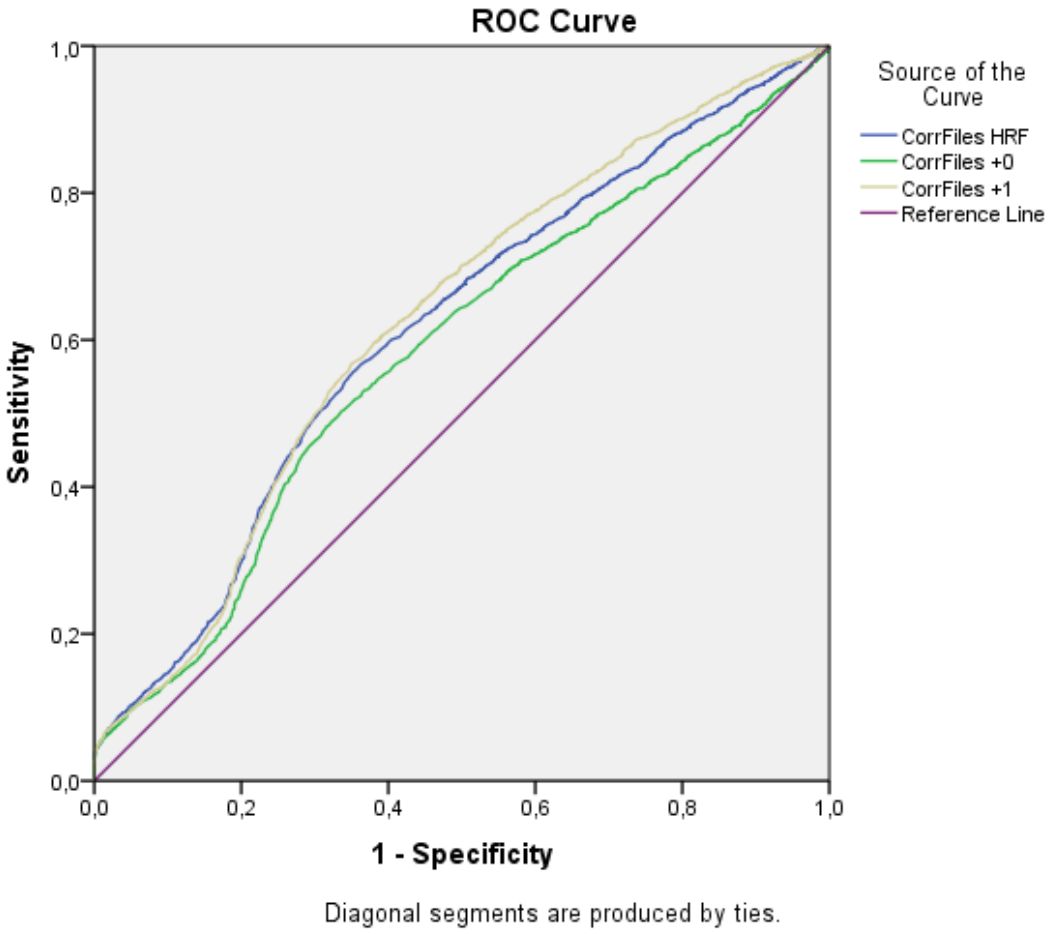

Figure 28. ROC curves of different reference functions. This figure shows ROC curves of the different reference functions over all data. The three lines represent the three reference functions “HRF”, “Shift+0” and “Shift +1”. Additionally a 45° reference line is drawn. “Shift +1” has the highest AUC, whereas “Shift +0” lies closest to the 45° reference line.

Figure 28 shows the outcome of applying the 3 different reference functions on all our ROC curve analysis data. “Shift +0” is the curve closest to the 45° reference line, whereas the “Shift +1” curve lies highest.

Table 23. Results of ROC curve analysis for each reference function.

HRF	AUC	.613
	Conf. Interv.	.601-.624
Shift +0	AUC	.583
	Conf. Interv.	.571-.594
Shift +1	AUC	0.625
	Conf. Interv.	.613-.636

AUCs of for the ROC curves of Figure 28 are shown in Table 23. As seen in Figure 28, “Shift +0” results in the lowest AUC and “Shift +1” has a slightly higher AUC than “HRF”.

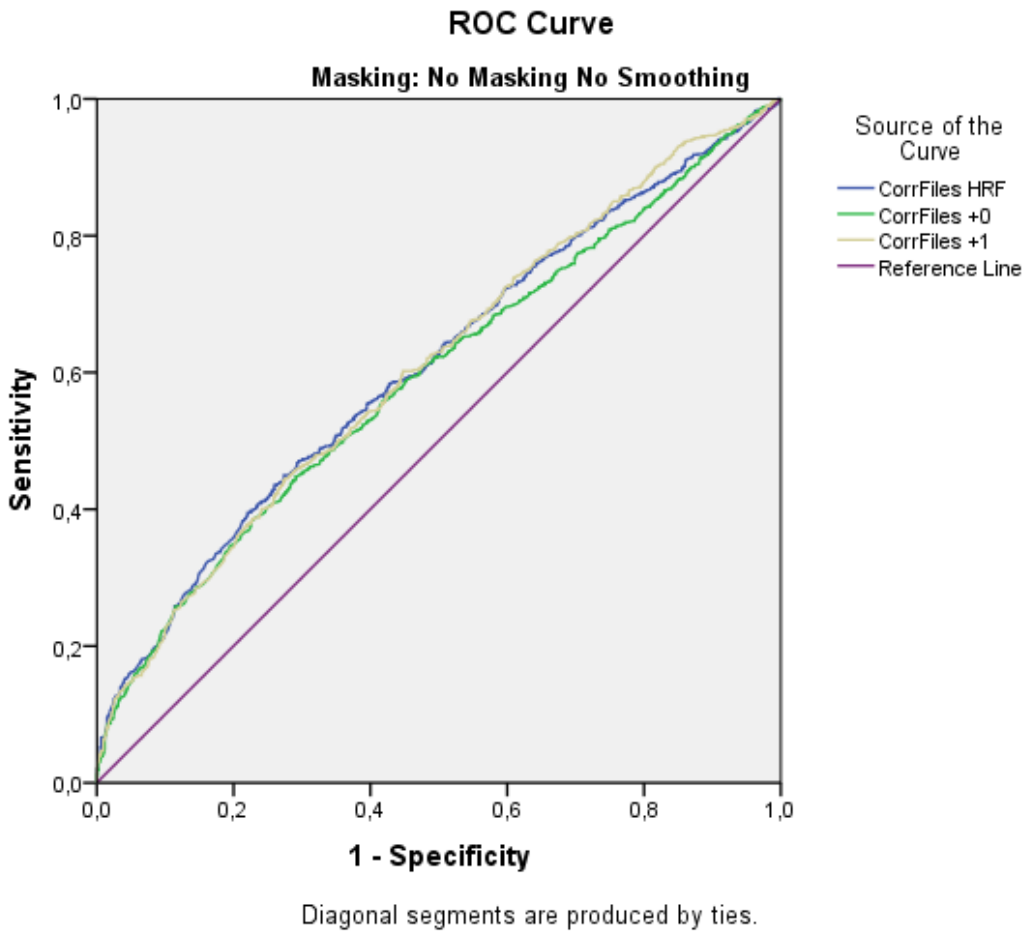

Figure 29. ROC curves of different reference functions for condition “No Mask, No Smoothing”. The three lines represent the three reference functions “HRF”, “Shift+0” and “Shift +1”. Additionally, a 45° reference line is drawn. The three lines are tangled, but “Shift +0” lies a bit lower in the upper range.

In Figure 29 the ROC curves for the different reference functions of the preprocessing condition “No Mask, No Smoothing” are shown. The 3 lines are tangled, no clear difference can be seen. Only in the higher range of sensitivity and 1-specificity the “Shift +0” curve is lower and closer to the 45° reference line.

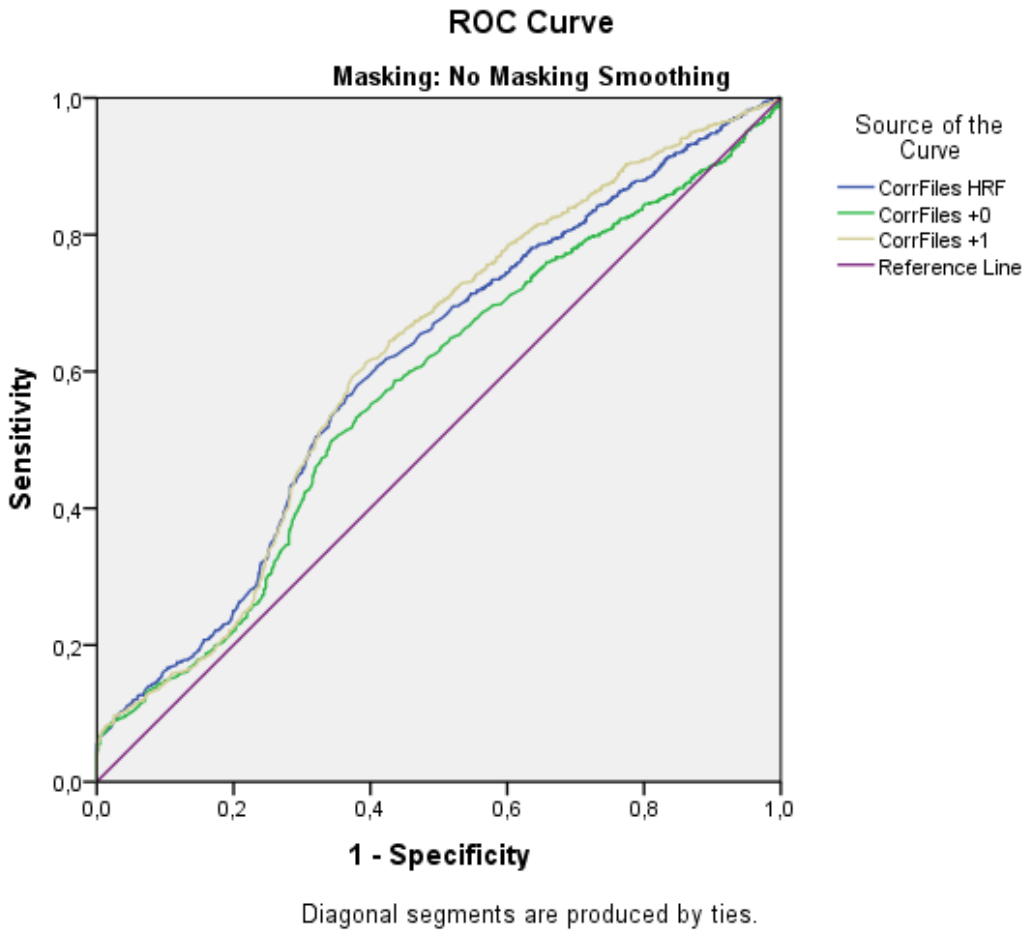

Figure 30. ROC curves of different reference functions for condition “No Mask, Smoothing”. The three lines represent the three reference functions “HRF”, “Shift+0” and “Shift +1”. Additionally, a 45° reference line is drawn. “Shift +1” lies highest above the 45° line, whereas “Shift +0” is located closest to it.

Results for the preprocessing condition “No Mask, Smoothing“ can be seen in Figure 30. Reference function “Shift +0” is nearest to the 45° reference line, especially in the highest range of sensitivity and 1 – specificity. Additionally, the curve for “Shift +1” seems to be better than the one for “HRF”.

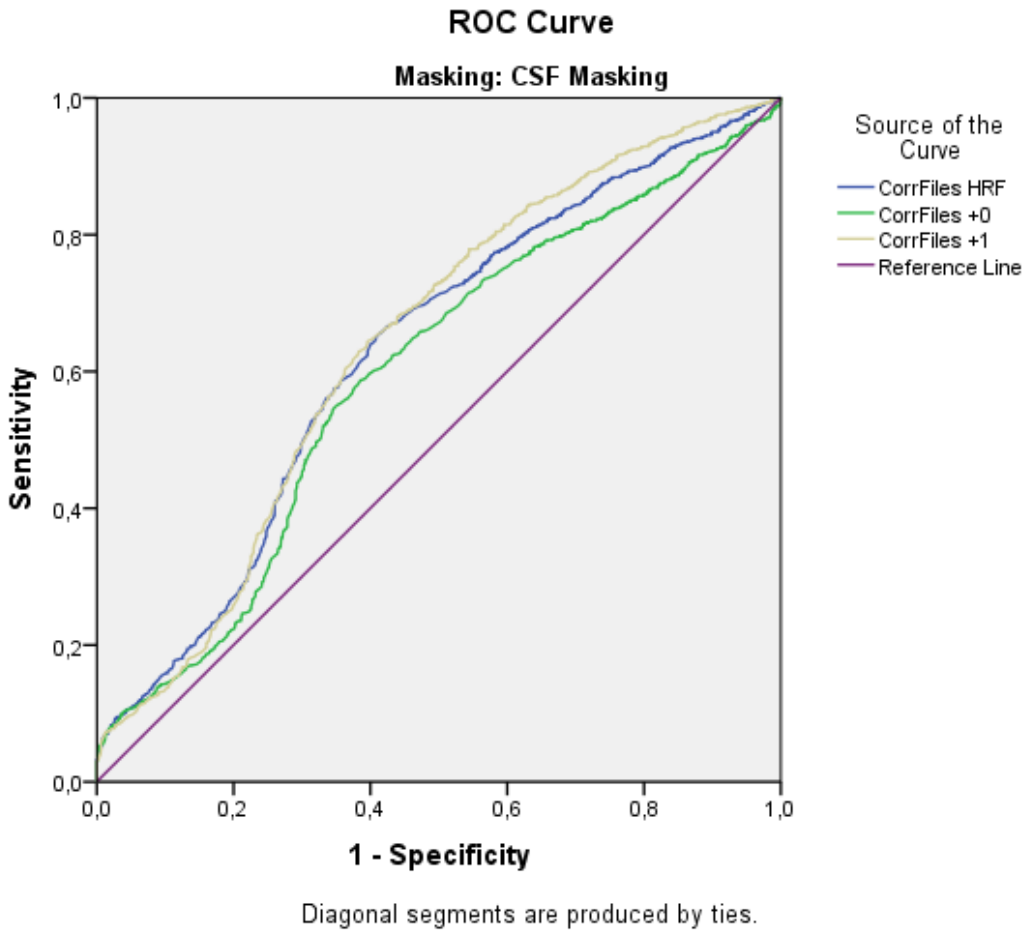

Figure 31. ROC curves of different reference functions for condition "CSF Mask, Smoothing". The three lines represent the three reference functions "HRF", "Shift+0" and "Shift +1". Additionally, a 45° reference line is drawn. "Shift +1" lies highest above the 45° line, whereas "Shift +0" is located closest to it.

Looking at the ROC curves for the different reference function of "CSF Mask, Smoothing", "Shift +0" again lies closest to the 45° reference line. Further, the "Shift +1" curve is higher than the HRF curve (see Figure 31).

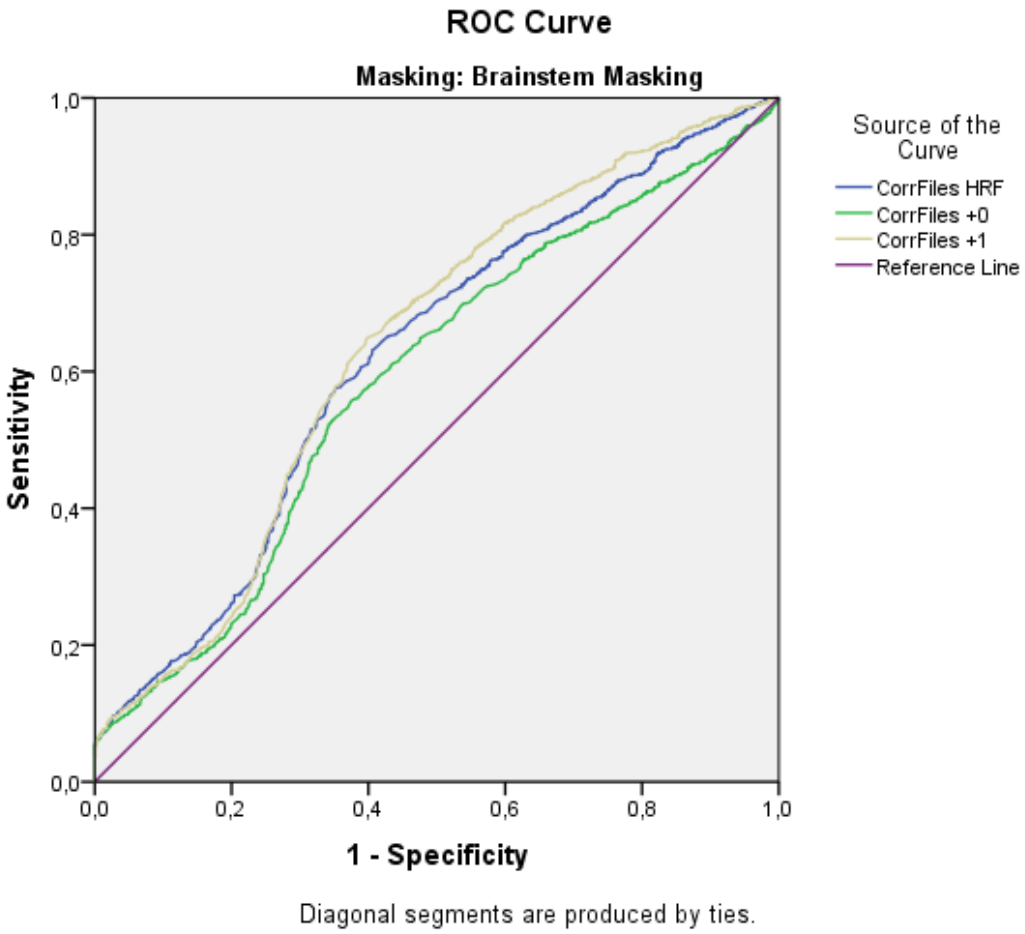

Figure 32. ROC curves of different reference functions for condition "BS Mask, Smoothing". The three lines represent the three reference functions "HRF", "Shift+0" and "Shift +1". Additionally, a 45° reference line is drawn. "Shift +1" lies highest above the 45° line, whereas "Shift +0" is located closest to it.

Figure 32 shows the influence of the different reference functions on the "BS Mask, Smoothing" condition. The same outcome as for "CSF Mask, Smoothing" can be seen. The curve presenting "Shift +1" is the highest, followed by the "HRF" curve and again, "Shift +0" lies closest to the 45° reference line.

Table 24. Results of ROC curve analysis for each reference function, separately for preprocessing conditions.

		HRF	Shift +0	Shift +1
No Mask, No Smoothing	AUC	.609	.594	.609
	Conf. Interv.	.585-.634	.569-.618	.585-.634
No Mask, Smoothing	AUC	.606	.571	.617
	Conf. Interv.	.583-.629	.548-.595	.593-.640
CSF Mask, Smoothing	AUC	.628	.595	.640
	Conf. Interv.	.605-.651	.572-.618	.617-.662
BS Mask, Smoothing	AUC	.621	.588	.635
	Conf. Interv.	.598-.644	.565-.612	.612-.658

In Table 24 AUCs of Figure 29 - Figure 32 are taken together. The reference function “Shift +0” yields the lowest AUCs in all cases. Therefore, it is the one with the least discriminatory power. Further, using “Shift +1” results in highest AUCs. “Shift +1” seems to be the best reference function to show activation of our data. This indicates, that brain stem BOLD signals include a relevant delay from task start to BOLD peak and therefore are best modelled with reference functions with delayed BOLD activity onset.

4. Discussion

4.1. Main findings

Comparing the sensitivity of risk maps of cortex and brainstem we found, that BOLD signals in motor cortex could be detected in 100% of all cases and BOLD signals in brainstem trigeminal motor nuclei in 36%. Further, the maximal correlation thresholds of cortex BOLD signals were most frequently between 0.5 and 0.8, whereas in brainstem trigeminal motor nuclei the maximal correlation thresholds with BOLD signal were overall lower, ranging from 0.2-0.65.

Concerning the influence of different preprocessing conditions on risk maps of brainstem, “No Mask, No Smoothing” led to activation in lowest correlation thresholds, for each subject separately as well as for all subjects together. When risk map data of all subjects together were evaluated, “No Mask, Smoothing” seemed to yield most activation in higher correlation thresholds, whereas separately for each subject this trend could not be seen anymore, showing that we found considerable between-subject variabilities. Regarding the influence of preprocessing conditions on brainstem data evaluated by ROC curve analysis, results showed highest AUCs with “CSF Mask, Smoothing” for all subjects together. When the analysis was done separately for each subject, no clear trend could be seen. ROC curve analysis was additionally done separately for each session of each subject. Here we could show that “No Mask, No Smoothing” yielded lowest AUCs, whereas “BS Mask, Smoothing” mostly led to highest AUCs, directly followed by “CSF Mask, Smoothing”. These outcomes also show that individual analyses may lead to different inferences than group analyses (Vandenbroucke et al., 2004).

Comparing the influence of the different reference functions on brainstem risk maps we found, that “Shift +1” led to the most active voxels in higher correlation thresholds. This result was supported by ROC curve analysis, where also “Shift +1” yielded the highest AUCs.

4.2. Comparing risk maps of cortex and brainstem

From literature we know that activations in brainstem nuclei are difficult to detect due to their small size and close location to large blood vessels and cerebrospinal fluid (Beissner, 2015; Schulte et al., 2016). Therefore, a decreased BOLD signal strength in brainstem nuclei compared to motor cortex can be expected. Indeed, we found that from 100% detectable activation in cortex, 36% were detected in brainstem. This demonstrates the relevant noise

contamination of BOLD signals within the brainstem. Further, brainstem activation was found in lower correlation thresholds. Correlation threshold provides information about the correlation of a voxel with our test parameters. If a voxel is activated during the task phases and not activated during resting phases, the performance of the voxel follows the test parameters, thus the correlation threshold is high, and vice versa. It has already been shown that motor cortex activation can be reliably detected using risk maps (Beisteiner et al., 2000). Also in our case motor cortex activation was detected in all cases and mostly in upper correlation thresholds between 0.5 and 0.8. The lower correlation thresholds and lower frequency of detectable BOLD signals in brainstem nuclei can have different reasons. Strong motion artefacts by adjacent blood vessels and CSF may occur (Beissner, 2015; Harvey et al., 2008). As risk map analysis works with minimized methodological assumptions, it doesn't account for motion parameters. Therefore, motion artefacts influence final risk map results and they include this experimental information. Note that in a clinical context, information about patient behavior and general run quality including noise level is important for individual clinical evaluations. It may result in the recommendation to repeat a clinical investigation instead of trying to define "residual activations" surviving "modeled corrections". With standard analysis procedures, regression and "correction" may result in a loss of true brain activation and loss of detectability of a compromised but physiological brain activation.

Further, brainstem nuclei are much smaller than cortical structures, leading to a decreased signal-to-noise ratio (SNR) (Beissner, 2015). We tried to account for this problem by using higher field strength (7 T) with high spatial and temporal resolution. However, there are also disadvantages coming with higher field strength. Field inhomogeneities increase, as well as physiological noise resulting from tissue motion and cardio-respiratory functions. Regarding field inhomogeneities, fieldmap correction was used (Matt et al., in submission; Robinson and Jovicich, 2011). Physiological noise corrections were not done, as already mentioned, as such corrections are not intended when risk maps are used.

4.3. Comparing influence of different preprocessing conditions on results of risk maps and ROC curves in brainstem scans

4.3.1. Benefit of smoothing

A clear result of our analyses is that our data benefit from spatial smoothing. Smoothing averages data values across nearby voxels. A weighted average value is calculated across neighboring voxels and each voxel is replaced by this value. Most commonly data is smoothed

by convolution with a three-dimensional Gaussian filter (or kernel). A Gaussian function with a specific width is used, where the width is defined as the Full Width at Half Maximum (FWHM). Smoothing with an appropriate kernel width can increase the signal-to-noise ratio (SNR), suppress noise and enhance task-related signals. On the other hand, spatial resolution is reduced. But gaining signal for activations across many voxels may outweigh the loss of smaller features. (Poldrack et al., 2011; Stippich, 2015) We used a 3 mm FWHM Gaussian kernel on our data. Our results showed that condition “No Mask, No Smoothing” led to fewest detectable BOLD signals. As no motion correction was done and because of the increased physiological noise in brainstem, a high amount of noise in our data was expected. Images evaluated with condition “No Mask, No Smoothing” may suffered most from this high amount of noise. As the other three conditions, in which smoothing has been done, yielded more detectable activation, smoothing may have eliminated some of the noise or enhanced task-related BOLD signals.

4.3.2. Benefit of masking for sensitivity and specificity

Using ROC curve analysis as an additional method to display the calculations of risk map technique, seemed to be beneficial. It showed which of the test conditions yielded high sensitivity and specificity and was good enough to be able to distinguish between true positive and false positive values, which in turn is useful for the question of validity of the results.

BS masks and CSF masks were drawn before smoothing, to avoid migration of unwanted signal (e.g. from CSF) into ROIs. It has already been shown that masking can improve detection of brainstem nuclei activation and may reduce artefacts (Beissner et al., 2014; Matt et al., in submission). When we looked at our results of ROC curve analysis for each subject and session, we found that the masking conditions (“CSF Mask, Smoothing” and “BS Mask, Smoothing”) mostly yielded highest AUCs (except for one case). Therefore, it seems that sensitivity and specificity can be improved by masking procedures. Interestingly, in case of subject BS05 “CSF Mask, Smoothing” led to very low AUCs, similar to “No Mask, No Smoothing”. CSF masks were created automatized by probabilistically assigning voxels to tissue classes like the CSF. BS masks on the other hand were drawn manually. Therefore, the CSF mask of subject BS05 may not fit perfectly and unwanted signal from the CSF may disturbed the evaluation. It may be beneficial to manually create BS masks, as they led to increased sensitivity and specificity in our ROC curve analyses. The benefit of BS masks on sensitivity and specificity has also been shown by Matt et al. (in submission).

4.4. Comparing influence of different reference functions on results of risk maps and ROC curves of brainstem scans

A very interesting finding was made concerning reference functions. The function “Shift +1” led to best results in risk map analysis and to highest AUCs in ROC curve analysis. Reference functions provide information about the hemodynamic response of the subjects to the task. Handwerker et al. (2004) found variations of HRFs across individuals and brain regions, which may lead to decreased t values, misestimates or false negative results in statistical analyses. Hemodynamic responses can particularly be altered because of pathologies. Brain tumors lead to an abnormal blood supply which can affect hemodynamic response (Wang et al., 2012). Neurovascular coupling (the relationship between neural activation and its changes in cerebral blood flow) can be altered in persons with cerebrovascular disease. Further, ageing is associated with changes in hemodynamic response, due to e.g. arteriosclerotic changes affecting elasticity of vessels. (D’Esposito et al., 2003) By comparing different reference functions, we wanted to know how the hemodynamic response is arranged in the brainstem. Interestingly “Shift +1”, our response function with 5 sec. (2 TR) latency, led to most detectable BOLD signals. This finding indicates a relevant delay from task start to BOLD peak in brainstem BOLD signals. Therefore they are best modelled with reference functions with delayed BOLD activity onset.

4.5. Risk map technique for brainstem fMRI

The initial question for this thesis was, how the standard risk map technique performs for detection of BOLD signals from brainstem scans. Risk maps work with minimized methodological assumptions, to stay as close to the original data as possible, which means that no first level statistics like motion correction or physiological nuisance regressors have been used. This is based on the difficulties to generate valid models for data correction, particularly in patients and individual subjects (Eklund et al., 2016). When Beisteiner et al. (2000) introduced the risk map technique as an improvement of presurgical patient evaluation in motor cortex fMRI scans, they used a correlation threshold of at least 0.5 ($r > 0.5$) to assess reliably activated voxels. Using risk map technique for our data we found that only in 36 % of all cases BOLD signals could be detected. When risk map results were separated for each subject, in 3 of 6 subjects activated voxels were found at a correlation threshold of at least 0.5 (BS03, BS05, BS06). Evaluating our risk map results, we found that spatial smoothing generally improved

BOLD signal strength. Further, masking conditions led to increased sensitivity and specificity (higher AUCs in ROC curve analyses).

These results indicate, that a standard risk map approach in brainstem fMRI may only work in selected subjects with minimized noise contamination. Experienced clinicians can easily evaluate the general signal to noise and contrast to noise situation, shown by a risk map result. Therefore, it may be expected that most often no adequate clinical report can be generated from standard brainstem risk map applications. The major issue here is the security for separating noise from physiological signals. Comparable to the electrophysiological evoked response recordings, clinical application of an fMRI analysis technique with minimized methodological assumptions requires more scan repetitions – a limiting factor for a patient investigation. Modelling noise as done with standard fMRI software packages might help in some cases, however, results always need to be checked from different perspectives since this models may lead to wrong inferences (Beisteiner 2017). Besides an increase of the amount of available data, there are also other options to improve risk map analysis of brain stem data, as described in the following.

4.6. Limitations

4.6.1. ROI selection and validation

The location of trigeminal motor nuclei (target ROIs) was based on coordinates published by Beissner et al. (2011). But as the individual anatomical location of trigeminal motor nuclei can vary, ROIs were drawn enlarged in respect of the expected volume of the nucleus. Nevertheless, it is possible that trigeminal motor activation lied outside the expected target ROI. Activation outside of ROI boundaries are not detected by the computer based ROC curve analysis. The visual analysis of risk maps could be beneficial in this case, as a person who evaluates the risk maps can include trigeminal activation even though it is not located exactly in the drawn ROI. On the other hand, visual analysis of risk maps requires an expert with profound clinical fMRI experience. The chosen control ROI does not contain motor or sensory nuclei, therefore no activation should appear during the tasks. If activated voxels are found in control ROIs, possible activated voxels in target ROIs are not counted as activated, as the image is considered as too noisy. However, if another region for control ROIs is used, results may be different.

4.6.2. Functional image quality

Our images were acquired using ultra high field strength (7 T). As already mentioned, higher field strength on the one hand increases SNR, on the other hand it comes with increased field inhomogeneities and increased physiological noise. With field map correction we tried to compensate for field inhomogeneities (Robinson and Jovicich, 2011), however, they probably cannot be corrected completely with this method. Physiological noise corrections cannot be used with risk map technique, but seemed promising when other evaluation methods were used (Beissner et al., 2014; Brooks et al., 2013; Harvey et al., 2008; Matt et al., in submission). Maybe it would be of advantage for risk map technique to measure brainstems with a 3 T fMRI. SNR would be lower, but the amount of noise would probably be reduced in advance.

A head coil was used for our scans, which led to some signal loss in caudal regions. A head-neck coil could be used instead, which could yield a higher SNR and better sensitivity for brainstem nuclei activation.

One idea of the risk map technique is to avoid artifacts in advance, instead of correcting for them. Therefore an individually prepared plaster helmet is usually done, to minimize head motion artefacts (Beisteiner, 2017; Beisteiner et al., 2000). This was not done in our case, but would maybe have led to less noise in advance.

4.7. Conclusion

The general situation of brainstem fMRI, which comes with many challenging aspects (e.g. small nuclei and physiological noise), is very difficult for all available data analysis techniques. For our evaluation technique with minimized methodological assumptions, we could show that signal strength benefits from smoothing and that sensitivity and specificity can be improved by masking conditions. Still, the results indicate, that a standard risk map approach in brainstem fMRI may only work with selected subjects / patients with minimized noise contamination. However, data improvement may be expected by using a head-neck coil, to yield a higher SNR, or by avoiding artifacts in advance by e.g. using dedicated head fixation.

5. References

Bear, M.F., Connors, B.W., Paradiso, M.A., 2007. Neuroscience. Lippincott Williams & Wilkins.

Beissner, F., 2015. Functional MRI of the Brainstem: Common Problems and their Solutions. Clin. Neuroradiol. 25, 251–257. https://doi.org/10.1007/s00062-015-0404-0

Beissner, F., Deichmann, R., Baudrexel, S., 2011. fMRI of the brainstem using dual-echo EPI. NeuroImage 55, 1593–1599. https://doi.org/10.1016/j.neuroimage.2011.01.042

Beissner, F., Schumann, A., Brunn, F., Eisenträger, D., Bär, K.-J., 2014. Advances in functional magnetic resonance imaging of the human brainstem. NeuroImage 86, 91–98. https://doi.org/10.1016/j.neuroimage.2013.07.081

Beisteiner, R., 2017. Can Functional Magnetic Resonance Imaging Generate Valid Clinical Neuroimaging Reports? Front. Neurol. 8. https://doi.org/10.3389/fneur.2017.00237

Beisteiner, R., Lanzenberger, R., Novak, K., Edward, V., Windischberger, C., Erdler, M., Cunnington, R., Gartus, A., Streibl, B., Moser, E., others, 2000. Improvement of presurgical patient evaluation by generation of functional magnetic resonance risk maps. Neurosci. Lett. 290, 13–16.

Braak, H., Ghebremedhin, E., Rüb, U., Bratzke, H., Del Tredici, K., 2004. Stages in the development of Parkinson's disease-related pathology. Cell Tissue Res. 318, 121–134. https://doi.org/10.1007/s00441-004-0956-9

Brooks, J.C.W., Faull, O.K., Pattinson, K.T.S., Jenkinson, M., 2013. Physiological Noise in Brainstem fMRI. Front. Hum. Neurosci. 7. https://doi.org/10.3389/fnhum.2013.00623

D'Esposito, M., Deouell, L.Y., Gazzaley, A., 2003. Alterations in the BOLD fMRI signal with ageing and disease: a challenge for neuroimaging. Nat. Rev. Neurosci. 4, 863–872. https://doi.org/10.1038/nrn1246

Eklund, A., Andersson, M., Josephson, C., Johannesson, M., Knutsson, H., 2012. Does parametric fMRI analysis with SPM yield valid results?—An empirical study of 1484 rest datasets. NeuroImage 61, 565–578. https://doi.org/10.1016/j.neuroimage.2012.03.093

Eklund, A., Nicholsd, T.E., Knutssona, H., 2016. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. PNAS 113, E4929.

Grzybowski, M., Younger, J.G., 1997. Statistical methodology: III. Receiver operating characteristic (ROC) curves. Acad. Emerg. Med. 4, 818–826.

Handwerker, D.A., Ollinger, J.M., D'Esposito, M., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. NeuroImage 21, 1639–1651. https://doi.org/10.1016/j.neuroimage.2003.11.029

Harvey, A.K., Pattinson, K.T.S., Brooks, J.C.W., Mayhew, S.D., Jenkinson, M., Wise, R.G., 2008. Brainstem functional magnetic resonance imaging: Disentangling signal from physiological noise. J. Magn. Reson. Imaging 28, 1337–1344. https://doi.org/10.1002/jmri.21623

Matt, E., Fischmeister, F., Ahmad, A., Robinson, S., Weber, A., Foki, T., Gizewski, E., Beisteiner, R., in submission. Improving sensitivity, specificity, and reproducibility of individual brainstem activation. Manuscr. Submitt. Publ.

Ogawa, S., Lee, T.-M., Kay, A.R., Tank, D.W., 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. 87, 9868–9872.

Park, S.H., Goo, J.M., Jo, C.-H., 2004. Receiver operating characteristic (ROC) curve: practical review for radiologists. Korean J. Radiol. 5, 11–18.

Poldrack, R.A., Mumford, J.A., Nichols, T.E., 2011. Handbook of functional MRI data analysis. Cambridge University Press, Cambridge New York Melbourne Madrid.

Robinson, S., Jovicich, J., 2011. B_0 mapping with multi-channel RF coils at high field: B_0 Mapping With Multielement RF Coils at High Field. Magn. Reson. Med. 66, 976–988. https://doi.org/10.1002/mrm.22879

Roessler, K., 2005. Evaluation of preoperative high magnetic field motor functional MRI (3 Tesla) in glioma patients by navigated electrocortical stimulation and postoperative outcome. J. Neurol. Neurosurg. Psychiatry 76, 1152–1157. https://doi.org/10.1136/jnnp.2004.050286

Schulte, L.H., Sprenger, C., May, A., 2016. Physiological brainstem mechanisms of trigeminal nociception: An fMRI study at 3T. NeuroImage 124, 518–525. https://doi.org/10.1016/j.neuroimage.2015.09.023

Stippich, C., 2015. Clinical functional MRI: presurgical functional neuroimaging. Springer.

Vandenbroucke, M.W.G., Goekoop, R., Duschek, E.J.J., Netelenbos, J.C., Kuijer, J.P.A., Barkhof, F., Scheltens, P., Rombouts, S., 2004. Interindividual differences of medial temporal lobe activation during encoding in an elderly population studied by fMRI. Neuroimage 21, 173–180.

Wang, L., Chen, D., Olson, J., Ali, S., Fan, T., Mao, H., 2012. Re-examine tumor-induced alterations in hemodynamic responses of BOLD fMRI: implications in presurgical brain mapping. Acta Radiol. 53, 802–811. https://doi.org/10.1258/ar.2012.120118

6. List of Figures

7. List of Tables

8. List of Abbreviations

AUC	Area under the curve
BOLD	Blood oxygenated level-dependend
BS	Brainstem
CSF	Cerebrospinal fluid
EPI	Echo planar imaging
fMRI	Functional Magnetic Resonance Imaging
FN	False negatives
FNR	False negative rate
FOV	Field-of-view
FP	False positives
FPR	False positive rate
Hb	Deoxygenated hemoglobin
HbO_2	Oxygenated hemoglobin
HRF	Hemodynamic Response Function
MGE	Multi-echo gradient echo
rCBF	Regional cerebral blood flow
ROC Curve	Receiver operating characteristic curve
Shift +0	Step function with 1 TR latency
Shift +1	Step function with 2 TR latency
SNR	Signal-to-noise ratio
TE	Echo time
TN	True negatives
TNR	True negative rate
TP	True positives

TPR	True positive rate
TR	Repetition time