Connection Science
Vol. 21, Nos. 2–3, June–September 2009, 161–175

Taylor & Francis
Taylor & Francis Group

GLM and SVM analyses of neural response to tonal and atonal stimuli: new techniques and a comparison

Simon Durrant[a]*, David R. Hardoon[b], André Brechmann[c], John Shawe-Taylor[b], Eduardo R. Miranda[a] and Henning Scheich[c]

[a]*Neuroscience and Aphasin Research Unit, University of Manchester, Manchester, M13 9PL, UK;* [b]*Department of Computer Science, University College London, London, WC1E 6BT, UK;* [c]*Speziallabor Nicht-Invasive Bildgebung, Leibniz Institute for Neurobiology, 39118 Magdeburg, Germany*

This paper gives both general linear model (GLM) and support vector machine (SVM) analyses of an experiment concerned with tonality in music. The two forms of analysis are both contrasted and used to complement each other, and a new technique employing the GLM as a pre-processing step for the SVM is presented. The SVM is given the task of classifying the stimulus conditions (tonal or atonal) on the basis of the blood oxygen level-dependent signal of novel data, and the prediction performance is evaluated. In addition, a more detailed assessment of the SVM performance is given in a comparison of the similarity in the identification of voxels relevant to the classification of the SVM and a GLM. A high level of similarity between SVM weight and GLM t-maps demonstrate that the SVM is successfully identifying relevant voxels, and it is this that allows it to perform well in the classification task in spite of very noisy data and stimuli that involve higher-order cognitive functions and considerably inter-subject variation in neural response.

Keywords: fMRI; support vector machine; general linear model; music cognition; machine learning

1. Introduction

Music has attracted a wide spread of research into how and why people create, perform and listen to music. In this paper, we focus on the study of tonality, which is a major topic in music theory (Piston and Devoto 1987) and music psychology (Krumhansl 1990), investigating tonal processing, and in particular the effect of distance along the circle-of-fifths (Shepard 1982) for key changes within the stimuli. Tonal melodies are complex structures, and cognitive processing reflects this structure (Narmour 1991). Analysing the neural activity associated with a melody is therefore particularly useful for determining which areas control higher-order sequence processing (Tervaniemi 2003).

Previous work in the area of tonality (Janata et al. 2002) made strong claims about the possibility of identifying a tonal map in the rostromedial prefrontal cortex. These results have not been reproduced to our knowledge. We set ourselves a somewhat complementary goal of testing whether differences between tonal and atonal stimuli can be detected, as well as differences correlating

*Corresponding author. Email: simon.durrant@manchester.ac.uk

ISSN 0954-0091 print/ISSN 1360-0494 online

DOI: 10.1080/09540090902733863
http://www.informaworld.com

with the distance along the circle-of-fifths as the stimuli change keys, neither of which were included in the Janata et al. (2002) study.

The central focus of our paper is in developing and testing a new form of analysis for fMRI data. Machine-learning methods have recently been applied to fMRI analysis (Carlson, Schrater and He 2003; Cox and Salvoy 2003; Wang, Hutchinson and Mitchell 2003; Mitchell et al. 2004; Davatzikos et al. 2005; Haynes and Rees 2005; LaConte, Strother, Cherkassky, Anderson and Hu 2005; Mourao-Miranda, Bokde, Born, Hampel and Stetter 2005; Kriegeskorte, Goebel and Bandettini 2006) to analyse the relationship between stimulus categories and fMRI responses. Higher-order cognitive effects are known to be difficult to detect in fMRI due to the presence of confounds and the nonlinear mapping between cognition and BOLD (blood oxygen level-dependent) response (the signal measured in fMRI), so fMRI data from a music task that focuses on such effects provide a significant challenge to machine-learning algorithms. A major focus of our paper is on the analysis of fMRI data with support vector machines (SVMs), both in combination with and in comparison to the more conventional and widely used fMRI analsyis approach employing general linear models (GLMs).

This paper is structured as follows. Section 2 discusses the materials and methods used in our experiment. This contains an elaboration on the experiment protocol, as well as providing information about the participants and acquisition of the data. Section 2 ends with a description of the pre-processing applied to the fMRI data. In Section 3, we first describe the new two-step procedure for combining the GLM and SVM and describe the haemodynamic response function (HRF) and how it affects the design matrices used in the analyses. We then go on to present the GLM analysis, the SVM analysis and a comparison between these two; this forms the main body of the paper. Finally, Section 4 brings the paper to a close with a brief summary of the analyses given and ways in which they can be extended.

2. Materials and methods

2.1. *Experimental design and stimuli*

The experiment was concerned with the tonality of short musical sequences. In particular, the focus of interest was the effect of relative tonality (the relationship between musical keys). Each stimulus consisted of 16 isochronous events lasting 500 ms each (with each stimulus therefore lasting 8 s), with no gaps in between; each event consisted of four simultaneous tones forming a consonant chord recognised in Western tonal music theory. The stimuli were created using the MIDI protocol, and rendered into audio files using a piano sound patch from the Roland Sound Canvas© digital samples. An example of the stimuli (in standard music notation) is shown in Figure 1. The stimuli were divided between tonal stimuli, which were designed to create a clear sense of key, and atonal stimuli[1] that were designed to create no clear sense of key, by the ordering of the chords, which were nevertheless equally consonant at the individual chord level in both types of stimuli. In order to verify this sense of key, the MIDI toolbox (Eerola and Toiviainen 2004) was used to test the stimuli with the Krumhansl–Schmuckler key-finding algorithm (Krumhansl 1990), This algorithm is based on findings from experimental psychology which connect the sense of key with the first-order distribution of tones within a melody, in a process of template matching. Enculturated listeners have a preference and expectation for some tones more than others in a given key, which allows key profiles to be constructed (Krumhansl and Kessler 1982); the profile for C major is shown in Figure 2 as an example. Measuring the inner product between the profile for each key and the distribution of tones within a sequence gives a vector of relative key likelihoods; the maximum element of this vector identifies the most likely key, and the strength of this element (scaled from 0 to 1) shows how well the piece matches

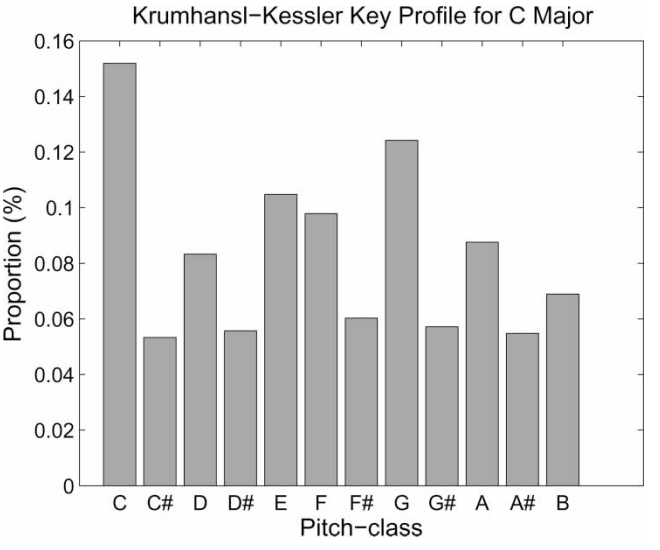

Figure 1. Krumhansl–Kessler (1982) key profile for C major, showing the preference for some tones over others.

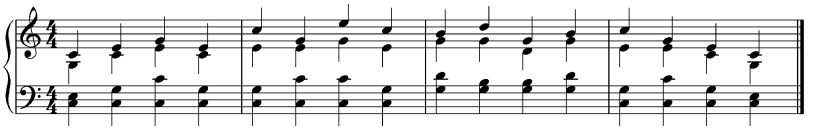

Figure 2. A single stimulus from the experiment (initial tonal condition).

the profile, which reflects how strongly the sequence sounds as though it is that particular key (a certainty rating). Empirical tests demonstrate that this algorithm reliably identifies the key that listeners actually perceive (Schmuckler and Tomovski 2005). Applying this algorithm to our stimuli, we obtained $\mu_{\mathrm{T}} = 0.93931$ and $\sigma_{\mathrm{T}} = 0.038562$ for the tonal stimuli and $\mu_{\mathrm{A}} = 0.65857$ and $\sigma_{\mathrm{A}} = 0.13196$ for the atonal stimuli, where μ is the mean of the strongest key certainty ratings (ranging between 0 and 1) and σ is the standard deviation of these ratings. The tonal and atonal stimuli were therefore expected to be reliably separable for participants. The experiment also contained a behavioural task to allow direct verification of this.

Altogether, 8 different tonal stimuli and 24 atonal stimuli were created. For a single run, stimuli were ordered into 24 groups of three stimuli with no gaps between stimuli or groups. The first stimulus in each group was always a tonal stimulus presented in the home key of C major, the second was always a tonal stimulus that could either be in the most distant key (around the circle-of-fifths) of F# major (first condition), the closest key (around the circle-of-fifths) of G major (second condition) or the same key of C major (third condition). The third stimulus in each group was always an atonal stimulus (fourth condition), which also reset the listener's sense of key (none of the participants possessed absolute pitch). As a result of the contiguity of groups, the first stimulus in each group followed the atonal stimulus in the previous group (except for the first group), which was therefore defined as the initial (fifth) condition. The first and second conditions therefore define changes from one key to another (distant or close). The third condition defines no change of key. The fourth condition defines no key present, and the fifth condition defines a change of no key back to a sense of key. The stimuli were ordered

such that all tonal stimuli were used an equal number of times, and conditions appeared in all permutations equally in order to control for order effects; the stimulus protocol is shown in Figure 3. This stringent order counterbalancing control was preferred to randomisation, and it also allowed us to determine the limits of within-subject variability across two consecutive runs within the experimental session without an additional possible contribution from order effects. The behavioural task for subjects was to click the left mouse button when they heard a change to a different key (conditions 1, 2 and 5), and right-click the mouse button when they heard a change from no key to a key (condition 4), in order to concentrate their attention on the tonal structure of the stimulus stream. The task was explained as clicking in response to a change (since non-musicians would not know what is meant by a key), and a short training session prior to scanning was used to ensure that subjects understood which type of change was being referred to. The behavioural results indicated that subjects clearly understood and were able to carry out the task.

2.2. *Subjects*

We tested 16 right-handed subjects with normal hearing (9 female, 7 male; aged 19–31) none of whom had received any formal musical education. All subjects gave written informed consent to the study, which was approved by the Ethics Committee of the University of Magdeburg.

2.3. *Data acquisition*

Functional magnetic resonance imaging data were acquired at the Leibniz Institute of Neurobiology (Magdeburg, Germany) on a Siemens Trio (Erlangen, Germany) 3T MRI scanner equipped with an eight-channel head coil. Functional volumes were collected using echo planar imaging with the following parameters: TE = 30 ms; TR = 2000 ms; interslice time: 62 ms; slice thickness: 3 mm; slice gap thickness: 0.3 mm; inplane resolution: 3 mm × 3 mm (giving 3 mm × 3 mm × 3 mm cubic isovoxels); number of slices: 32; FA: 80°; FOV: 192 mm × 192 mm; matrix size: 64 × 64. Stimulus delivery and scanning coordination were controlled with the Presentation© software (Neurobehavioural Systems Inc., Albany, USA) using a custom-written script. The perceived scanner noise was attenuated by earplugs (24 dB) and ear muffs (20–30 dB) in which MRI-compatible electrodynamic headphones were integrated (Baumgart et al. 1998). The loudness of the stimuli was individually adjusted to a comfortable level. Each stimulus block lasted 8 s (4 volumes) and was immediately followed by the next stimulus block. Two experimental runs were carried out during the session, with 20 s (10 volumes) silence before each run, and after the final run, to provide a baseline condition. Altogether, each session therefore consisted of 606 functional volumes, as well as anatomical data collection and dummy runs for scanner alignment. Subjects were also given an initial scan-free practice period on stimuli not used in the functional data collection in order to ensure that they understood the task.

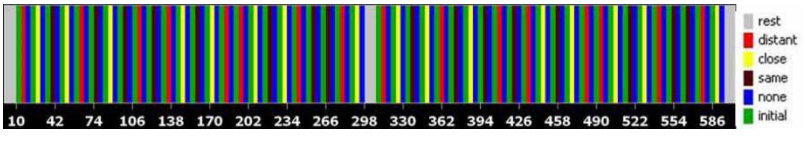

Figure 3. Stimulus protocol for the experiment, showing the order and timing of the different experiment conditions.

2.4. *Pre-processing*

Functional data from each participant in their original brain space were co-registered with in-plane anatomical data and pre-processed with 3D motion correction using trilinear interpolation, slice scan time correction using sinc interpolation, linear trend removal, high-pass temporal filtering with a cut-off of three cycles throughout the time course, equivalent to 0.00236593 Hz and Gaussian spatial smoothing with an FWHM of 4 mm. A correlation analysis was performed in the original data space following pre-processing to verify its effectiveness. The data were then co-registered with a high-quality anatomical data set and transformed into AC-PC space using using cubic spline interpolation and subsequently into Talairach space with trilinear interpolation, providing a standard brain space for all participants to allow group-level analysis. Finally, a grey matter mask was estimated individually for each participant using intensity analysis with cortical probability maps, and these masks were combined on an inclusive basis with a logical 'OR' function so that all voxels that were designated grey matter from one or more participants remained. All other voxels were screened out, and not included in further analysis; this process reduced the number of voxels per participant from 131,072 to 54,887. All pre-processing was done using BrainVoyager QX© (Brain Innovation B.V., Maastricht, The Netherlands). These pre-processed data formed the basis for all subsequent analyses.

3. Experimental results and analysis

3.1. *Overall approach*

In order to develop and evaluate an effective machine-learning approach to the analysis of fMRI data, we focused our attention on one particular aspect of the experiment: tonal *vs*. atonal stimuli. This is captured by experimental conditions 4 (atonal) and 5 (initial tonal), so our subsequent analyses here are concerned entirely with these two experimental conditions. Two paths of analysis were taken: a conventional approach using statistical parametric maps (Friston, Ashburner, Kiebel, Nichols and Penny 2006) in a mass univariate GLM analysis and a machine-learning approach using an SVM. There is an important difference between these approaches: the GLM is a regression analysis and seeks to identify which voxels (volumetric pixels; the individual units of volume in an fMRI data set) are most relevant given a set of predictor variables that specify which condition is active at a given time point, whereas the SVM is a classification analysis and seeks to identify the predictor variables (i.e. which condition is active at a given time point) for a set of unseen test data after first being trained on other data; this is effectively Bayesian inference. As such, these two approaches may be seen as complementary. However, they also have a common point of reference, which is the identification of the relative importance of each voxel to distinguish between experimental conditions. This provides a means to compare the performance of these two approaches, and this comparison is the ultimate goal of our analysis here.

BOLD response in fMRI for higher-order cognitive tasks, including the processing of tonality in music, is notoriously noisy data. In addition, the amount of data (in this experiment 64×64 voxels per slice, 32 slices per volume, 606 volumes per participant, 16 partipants, which gives 1.2709×10^9 data points altogether), while providing ample amounts of training data, poses a serious challenge to machine-learning algorithms that have to seek for an optimal solution within a very large search space. It is clear from conventional analyses that only a selection of relatively small brain areas are relevant for any given task, and by limiting analyses to these areas, the solution search can be made a lot more efficient. The identification of relevant active areas should be as inclusive as reasonably possible (i.e. careful not to exclude areas that might be of interest), and as such a fixed-effects group level analysis is appropriate (see Section 3.3 for more details),

which includes all voxels that are significant for the subject group as a whole, ignoring the variance between subjects (thus making it much more inclusive than random effects analysis which also accounts for between-subject variance). This forms the basis of our overall two-step analytical procedure, which is as follows.

(1) Perform second-level (group) fixed-effects GLM analysis and create mask based on active voxels when contrasting conditions 4 (atonal) and 5 (initial tonal), for 15 participants leaving one out (for each combination of participants, hence 16 masks altogether).
(2a) Perform first-level (individual) GLM analysis for one participant using the mask from the other 15 participants developed in the first stage (as described in Section 3.3). This is repeated for each participant in turn.
(2b) Perform SVM analysis, training on 15 participants using the mask from the other 15 participants developed in the first stage and testing on the remaining participant (a leave-one-out strategy) (as described in Section 3.4). This is repeated for each participant in turn.

Finally, the results of the two alternative second-stage analyses (GLM and SVM) are compared, as described in Section 3.5.

3.2. *Haemodynamic response, lag and design matrices*

The five experimental conditions, plus a sixth baseline condition, can be treated as dummy variables in time with a 1 representing the condition being active for a given time point and a 0 otherwise. These can be organised into a 606×6 design matrix, which provides the core experiment model for both the GLM and SVM analyses. Of the 606 time points, 192 have condition 4 active, 192 have condition 5 active and the remainder are divided between the other three conditions (which we are not concerned with here).

In order for these to act as predictor variables, however, the lag between the stimulus presentation, and the neural BOLD response, has to be taken into account. There are essentially two related ways to do this. One is simply to shift the variables in time, encoding a lag in the design matrix. An advantage of this approach is that the essential form of the design matrix, consisting of 0 and 1 s only and with orthogonal experimental conditions, remains intact. This is important for an SVM engaged in binary classification of the two conditions, where a 'correct' answer of only one active condition per time point is required for training and performance measurement. However, this leaves open the question of how long the lag should be. An alternative approach widely adopted in conventional fMRI analysis is to use an HRF modelled from gamma functions with parameters determined by biophysical data. Each column in the design matrix is convolved with this function to provide a set of predictor variables for the expected neural response to the different stimulus conditions. The HRF used in the design matrix for the GLM is shown in Figure 4. Figure 5 shows the design matrices for simple lag 0, simple lag 3 and HRF convolution, respectively (first 30 volumes displayed only for space reasons).

3.3. *Conventional GLM analysis*

The conventional analysis was performed with a GLM using a combination of BrainVoyager QX 1.9© (Brain Innovation B.V.) and MATLAB 7.3© (The Mathworks, Cambridge, MA, USA). A GLM is a widely used statistical method that is a generalisation of both analysis of variance and multiple linear regression and subsumes other popular techniques such as t-tests. The central equation can be written in the matrix form as $\mathbf{Y} = \mathbf{X}\beta + \epsilon$, where $\mathbf{Y}$ is the measured signal (an $M \times N$ matrix which can be interpreted as N dependent variables each containing M data records), $\mathbf{X}$ is an $M \times Q$ design matrix of predictor (independent) variables (see Section 3.2 for

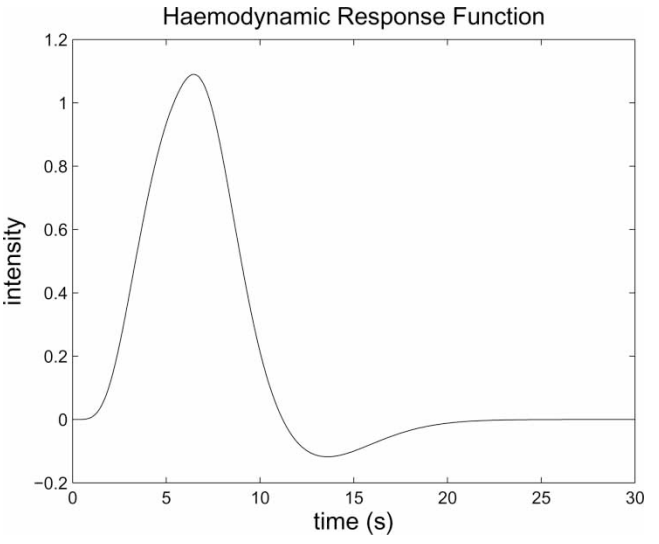

Figure 4. HRF, showing the typical time course of the BOLD response measured in fMRI. The function shown here is the one with which the design matrix for the GLM analysis was convolved.

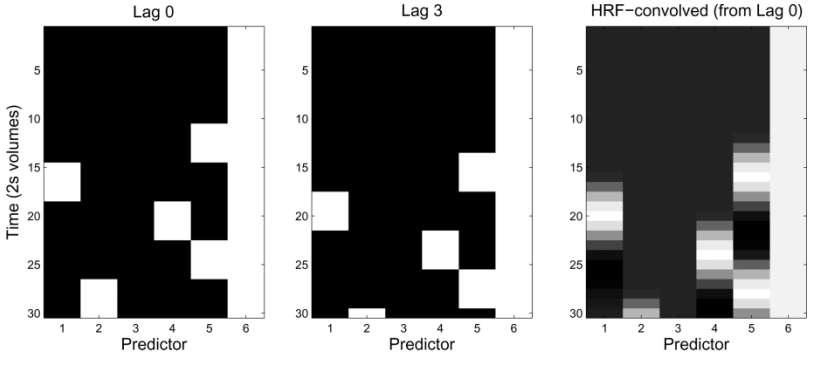

Figure 5. Stimulus protocol for the experiment, showing the order and timing of the different experiment conditions.

details of the design matrices used in our analysis), β are a $Q * N$ set of linear coefficients to be estimated and ϵ represents the residual error. The coefficients are estimated using ordinary least squares (to minimise the residual error), such that $\widehat{\beta} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} = \mathbf{X}^{+}\mathbf{Y}$ where $\mathbf{X}^{+}$ is the Moore–Penrose pseudo-inverse. The estimate for each β coefficient can be considered the mean of a distribution of possible values for that coefficient, which represents the most probable value (the one that gives the lowest residual error). Each β coefficient also therefore has an associated variance (the variance of the estimate distribution, which is also reflected in the variance of the residuals ϵ for the variable in question); the larger this variance, the less well the coefficient value fits for all the data records of the variable the coefficient represents, and therefore the less statistically significant the estimate is likely to be. In other words, the coefficient variance (strictly, standard deviation), relative to the size of the coefficient value, gives us a direct measure of the explanatory power of the coefficient. This is a t-statistic, and is given by $t_{\beta_i} = \beta_i / \sqrt{\sigma^2_{\beta_i}}$ for a given coefficient β_i.

For conventional fMRI analysis, each voxel is treated as an independent time series, which means that a mass univariate approach is employed, with a separate GLM estimate of β coefficients

for each voxel. The number of β coefficients is dependent on the the number of explanatory variables (columns) in the design matrix. In addition to estimating a t-statistic for each β coefficient individually, the significance of a combination of variables for a given voxel can be examined by adding or subtracting β coefficients. This leads to the concept of a contrast vector, which is simply a set of coefficients (typically 1, 0 or −1) that are applied to the β coefficients (and their standard deviations), to give a t-statistic which reflects the significance of a combination of variables or a difference of variables for a given voxel. Applying the same contrast vector for all voxels gives a t-map, also known as a statistical parametric map (Friston, Holmes and Worsley 1995), which is the core of GLM fMRI analysis. T-maps give a clear indication of which voxels are relevant, and to what extent, for a given combination of experimental conditions. Our present study exclusively uses the contrast vector [0 0 0 −1 1], which is the contrast tonal *vs*. atonal where a tonal stimulus always follows an atonal one, and vice versa, giving the most equitable comparison.

As described in Section 3.1, the GLM analysis was employed in two different ways. A second-level (group) analysis was performed in stage one, which first involves computing a GLM (per voxel) for each individual participant, and then taking the statistics from these forward into a second-level analysis examining the difference between subjects (this is also known as the summary statistic approach and is much more efficient than computing second-level GLMs for all voxels and participants simultaneously, while giving identical results; see Boser, Guyon and Vapnik (1999) for details). An important distinction in second-level analyses is between fixed-effects and random-effects analyses. Both types of analyses identify voxels that are significantly active across the group of participants in the experiment, but random-effects analysis also considers the variability of the activation of each significant voxel across the participants, only retaining voxels on which participants are significantly in agreement (have low variance across the group, as well as low variance within each participant). In our present study, we are using group analysis to identify relevant voxels for further analysis of different types. As such, we use the more inclusive fixed-effects analysis to identify relevant voxels. T-maps were created as described above, after which Bonferroni correction for multiple comparisons was applied (to minimise the occurrence of false positives), and finally a minimum cluster size of 50 mm^2 was imposed. Figure 6 shows examples of these correct t-maps for some slices when one given participant is left out, and a similar example when another participant is left out. Very similar areas of activation can be seen in both cases: bilateral precentral gyrus (Brodmann's area 4), as well as the left medial frontal

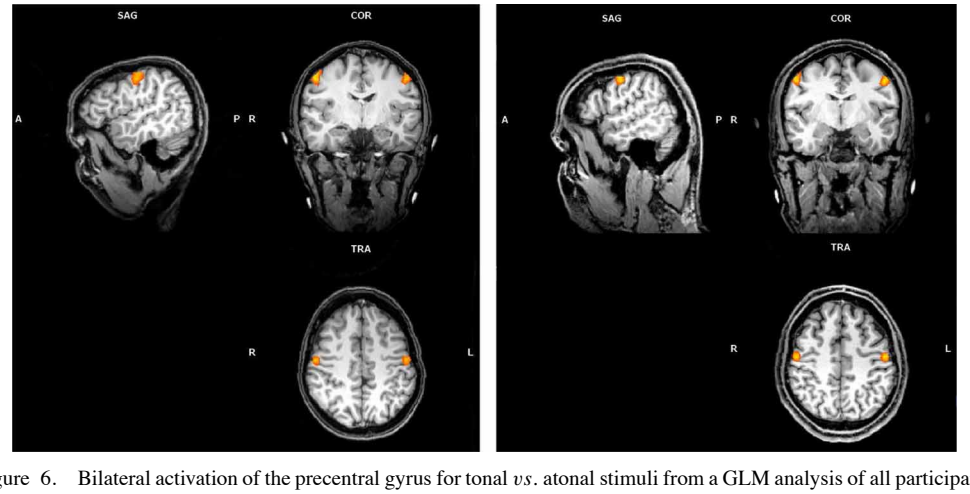

Figure 6. Bilateral activation of the precentral gyrus for tonal *vs*. atonal stimuli from a GLM analysis of all participants except z178 (top) and iq53 (bottom), shown against the anatomical scan for those two participants, respectively.

gyrus (Brodmann's area 6; not shown). Masks were produced using this procedure leaving out each one of the 16 participants in turn.

Following the creation of the leave-one-out masks in stage one, the second stage of the GLM analysis involved analysing each individual with a GLM with voxels identified by the leave-one-out analysis for that participant. That is, voxels that were identified as being significantly involved in the tonal *vs*. atonal analysis for the other 15 participants were then tested for the remaining participant using a GLM, in order to see whether or not they were also relevant for this participant. The results of the GLM analysis are shown in Table 1. The first column identifies the test participant left out of the group analysis; the number of voxels identified as active for all except one individual participant is shown in the second column; the number of those voxels that are active for the participant as well is shown in the third column and the average t-statistic for all of the voxels of the the second column is shown in the fourth column. These results highlight the fact that there are both significant variations between participants and also that voxels identified as significantly involved in the processing of tonality for a group of participants are also frequently significantly involved for a participant not in the original group; in other words, that some generalisation capability from one group of participants to another participant seems to be possible. It is this fact that motivates the machine-learning approach of the SVM, described in the next section.

3.4. *Machine-learning SVM analysis*

In order to test the ability of machine-learning analysis to detect patterns in fMRI data relating to higher order cognitive tasks, we adopt a machine-learning framework of SVMs (Vapnik 1995; Cristianini and Shawe-Taylor 2000) for a leave-subject-out analysis, i.e. we learn a discriminatory task on the combined data of 15 subjects and test on the remaining subject. SVMs are kernel-based methods that find functions of the data that facilitate classification. They are derived from statistical learning theory (Vapnik 1995) and have emerged as powerful tools for statistical pattern recognition (Boser, Guyon and Vapnik 1992). In the linear formulation, an SVM finds, during the training phase, the hyperplane that separates the examples in the input space according to their class labels. The SVM classifier is trained by providing examples of the form $(\mathbf{x}_a, y)$, where $\mathbf{x}_a$ represents an input and y its class label. Once the decision function has been learned from the training data, it can be used to predict the class of a new test example. In the present study, $\mathbf{x}_a$

Table 1. Results of GLM analysis, showing the number of voxels identified as important in the tonal *vs*. atonal contrast, for the group (leaving out the participant in question) and the participant (from the voxels previously identified), and also showing the average t-statistic for the voxels.

Participant identifier	All except voxels	Participant voxels	Mean t-statistic
au70	213	0	1.1224
di55	122	37	2.2076
dw43	107	47	2.0999
ed32	144	22	1.3644
fi68	142	32	1.817
fv95	147	0	1.0248
ic12	149	18	1.2468
iq53	181	43	1.5525
rn39	158	11	1.4246
sc50	198	9	1.3198
ty16	148	20	1.3627
ue70	135	0	0.93331
us10	110	53	2.3976
uv15	188	13	1.3062
xp38	137	0	1.2127
z178	124	43	2.0716

represents an fMRI observation and y is the task performed ($y = 1$ for task 1 (tonal) and $y = -1$ for task 2 (atonal)). For a detailed description of SVMs, see Cristianini and Shawe-Taylor (2000). We used a linear kernel SVM that allows direct extraction of the weight vector as an image (i.e. the discriminating spatial pattern). A parameter C that controls the trade-off between training errors and smoothness was fixed at $C = 1$ for all cases (the default value).[2]

Initial tests of the SVM model on the whole gave near-chance results. From these it was clear that even when masking to grey-matter only (the spatial-temporal information), to reduce the dimensionality of the data (from 131,072 to 54,887 voxels per participant), and the subtraction of the baseline scans (Mourao-Miranda, Bokde, Born, Hampel and Stetter 2005) to further reduce noise, the finding of a hyperplane that would accurately distinguish between the two tasks was extremely difficult for the learning algorithm. We hypothesised that in this type of complex stimuli, a further more precise dimensionality reduction is required. As such, we developed the strategy to use the GLM analysis on the leave-subject-out routine to generate this new mask, i.e. the GLM analysis is performed *only* on the training subjects, as described in Section 3.3. The mask is used to considerably reduce the dimensionality of the fMRI data used in the SVM learning procedure. In addition to this, we took an average over the four volumes that made up each stimulus block in our experiment design (because each stimulus lasted 8 s and each volume was 2 s in duration), giving us one new data point for each stimulus block. The result of this dimensionality reduction is that for each participant we originally had 54,887 voxels of interest, each with 384 volumes (for the two conditions of interest here), and after reduction, we have 100–200 voxels of interest (depending on which participant is left out; the exact figures are shown in the second column of Table 1) with 96 volumes each. In addition, we further pre-processed the data by subtracting from each subject the mean of the volumes representing the control (silence) condition in order to remove baseline noise from the data.

The exact procedure employed is detailed in the following pseudo-code.

```
For each leave-subject-out procedure:
1. We use the GLM analysis to produce a voxel mask

2. For each training subject;
2.1 Training subject is loaded and zero meaned (inter subject)
2.2 Mean value computed from the control blocks
2.3 Mean value computed for each block condition
2.4 Subtract 2.2 from 2.3

3. Test subject loaded and zero meaned (inter subject)
3.1 Mean value computed for each block condition
3.2 Subtract 2.2 from 3.1

4. The training data are zero meaned (across subjects)
4.1 The testing data are zero meaned using 4

5. SVM Training/Testing procedure.
```

We then run four sets of experiments in which we varied the lag encoded in the design matrix from zero to three (see Section 3.2), in order to evaluate the effect of the lag and to ensure that our tests included an optimal biophysically plausible lag. Table 2 summarises the SVM results, and we are able to substantiate our hypothesis that the use of the GLM mask does indeed improve the results across subjects, with the lag of one volume producing the lowest individual error for any one participant.

Table 2. SVM error values from the leave-subject-out experimentation for different haemodynamic lags (0–3).

Subject	Lag 0	Lag 1	Lag 2	Lag 3
au70	0.3646	0.7188	0.6875	0.5938
di55	0.5625	0.3958	0.4479	0.5417
dw43	0.4167	0.3646	**0.3333**	0.4688
ed32	0.3854	0.4271	0.3750	0.3750
fi68	0.4167	0.2917	0.3438	**0.3646**
fv95	0.4896	0.4479	0.4583	0.4375
ic12	0.4167	0.3542	0.3646	0.5000
iq53	0.3854	0.4896	0.4688	0.4583
rn39	0.5833	0.6771	0.5833	0.4896
sc50	0.4375	0.3958	**0.3333**	0.4062
ty16	0.3750	**0.2500**	0.3542	0.4271
ue70	0.4167	0.4375	0.5000	0.5000
us10	**0.3229**	0.3542	0.3438	0.3958
uv15	0.5625	0.5208	0.5312	0.5417
xp38	**0.3229**	0.4167	0.5312	0.5729
z178	0.3646	0.3750	0.3958	0.4062

Note: The lowest error values as well as the subjects who achieved them are shown in bold.

We can see from the SVM results shown in Table 2, as also seen in the GLM results (Table 1), that performance differs considerably between individuals. This is not surprising given the nature of the experiment with musical stimuli. Any significant agreement between individuals (which is implied by any above-chance performance in a leave-one-out scenario), and any significant agreement between methods, on such stimuli, can be regarded as indicative of a successful technique. In the following section, we compare the SVM and GLM results to see whether or not our current approach fulfils this promise.

3.5. *Comparison between GLM and SVM results*

As the GLM analysis in the second stage of our overall procedure performs regression, while the SVM analysis performs classification, it may at first glance seem impossible to directly compare their performance. However, we can in fact go beyond the relatively simple comparison of correct percentages for each method and compare the similarity of their performance at the voxel level. The use of a linear kernel in the SVM means that as well as getting performance measures in terms of correct predictions out of the model, we also get a set of weights that can be directly interpreted as the relevance of each individual voxel in distinguishing between tonal and atonal stimuli. In other words, we have a way of identifying which voxels are most useful in the classification of the different stimuli, and a linear measure as to what extent. For the GLM analysis, we saw earlier that for each voxel we have a t-statistic from the tonal *vs.* atonal contrast, which also tells us the relevance of each voxel, and also gives us a linear measure as to what extent. Our approach is therefore to directly compare these results by measuring the correlation between them, and as well having t-maps from the GLM, we therefore also have pseudo-t-maps from the SVM (rescaled according to the mean and standard deviation of the known 15 participants).

The results of our direct comparison are shown in Tables 3 and 4. These tables show the results for each of lags 0–3 inclusive, in groups of four columns for each lag, plus an initial participant identifier column. The first column in each group gives the correct predictions proportion from the SVM; these results are simply 1-error given in Table 2. The correct proportion is simply the number of correct predictions out of the 96 made for each participant. As this is a two-forced choice situation (tonal or atonal), chance level is 48 correct (0.5 correct proportion). A Pearson's chi-square test at the $p < 0.05$ level with 1 df can be used to determine whether the results differ

Table 3. Comparison of the SVM and GLM results (for lags 0 and 1), showing the proportion of correct results, a statistical significance flag for those results, the mean correlation between the SVM weights and GLM t-statistics at the voxel level and a statistical significance flag for that correlation.

	Lag 0				Lag 1			
Participant	SVM correct	Significance	Correlation with GLM	Significance	SVM correct	Significance	Correlation with GLM	Significance
au70	0.635417	1	−0.006517	0	0.28125	−1	0.03496	0
di55	0.4375	0	0.006987	0	0.604167	1	0.18857	1
dw43	0.583333	0	0.445105	1	0.635417	1	0.459884	1
ed32	0.614583	1	0.379157	1	0.572917	0	0.407643	1
fi68	0.583333	0	0.233451	1	0.708333	1	0.215785	1
fv95	0.510417	0	0.122586	0	0.552083	0	0.234174	1
ic12	0.583333	0	0.408898	1	0.645833	1	0.402016	1
iq53	0.614583	1	0.217314	1	0.510417	0	0.226806	1
rn39	0.416667	0	0.069144	0	0.322917	−1	−0.108302	0
sc50	0.5625	0	0.347325	1	0.604167	1	0.425107	1
ty16	0.625	1	0.327322	1	0.75	1	0.274961	1
ue70	0.583333	0	0.597773	1	0.5625	0	0.523077	1
us10	0.677083	1	0.750951	1	0.645833	1	0.775507	1
uv15	0.4375	0	−0.078438	0	0.479167	0	0.001754	0
xp38	0.677083	1	0.525801	1	0.583333	0	0.580896	1
z178	0.635417	1	0.487857	1	0.625	1	0.539654	1

Table 4. Comparison of the SVM and GLM results (for lags 2 and 3), showing the proportion of correct results, a statistical significance flag for those results, the mean correlation between the SVM weights and GLM t-statistics at the voxel level and a statistical significance flag for that correlation.

	Lag 2				Lag 3			
Participant	SVM correct	Significance	Correlation with GLM	Significance	SVM correct	Significance	Correlation with GLM	Significance
au70	0.3125	−1	0.041538	0	0.40625	0	−0.055864	0
di55	0.552083	0	0.184603	1	0.458333	0	0.124819	0
dw43	0.666667	1	0.465121	1	0.53125	0	0.344471	1
ed32	0.625	1	0.4015	1	0.625	1	0.268042	1
fi68	0.65625	1	0.17821	1	0.635417	1	0.137552	0
fv95	0.541667	0	0.277066	1	0.5625	0	0.345436	1
ic12	0.635417	1	0.410925	1	0.5	0	0.256487	1
iq53	0.53125	0	0.256479	1	0.541667	0	0.242307	1
rn39	0.416667	0	−0.229115	−1	0.510417	0	−0.338237	−1
sc50	0.666667	1	0.383195	1	0.59375	0	0.255643	1
ty16	0.645833	1	0.302665	1	0.572917	0	0.139682	0
ue70	0.5	0	0.38429	1	0.5	0	0.243081	1
us10	0.65625	1	0.776727	1	0.604167	1	0.630869	1
uv15	0.46875	0	0.032919	0	0.458333	0	0.046247	0
xp38	0.46875	0	0.576862	1	0.427083	0	0.569061	1
z178	0.604167	1	0.618461	1	0.59375	0	0.572237	1

significantly from the chance level; a flag indicating significant results is shown in column 2 (+1 is significantly above chance, −1 is significantly below chance, 0 is not significantly different from chance). The third column shows the mean correlation between the SVM weights and the GLM t-statistics, and the fourth column shows whether the correlation is statistically significant or not (+1 is a significantly positive correlation, −1 is a significantly negative correlation, 0 is no significant correlation).

It is clear from these results that there is a very significant amount of agreement between the two methods, with only two significant negative correlations and 47 significant positive correlations out of a total of 64. In addition, for all but two of the cases where the SVM has prediction

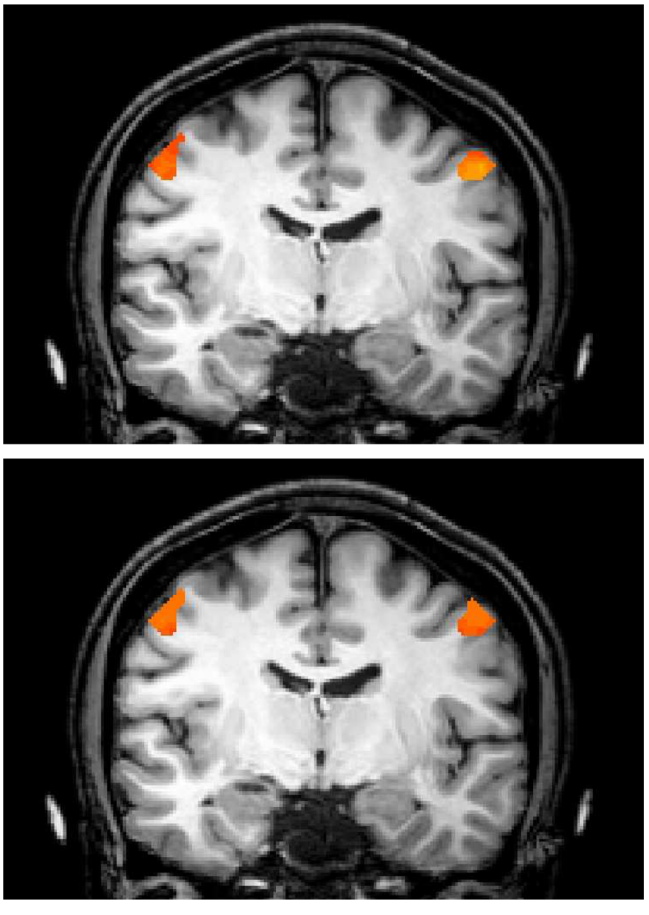

Figure 7. Activation maps for participant us10. *Top*: t-map from the GLM analysis. *Bottom*: SVM weights for lag 2.

significantly above the chance level, there is also a significant positive correlation between the SVM weights and the GLM t-statistics. An example of the similarity in the SVM weights and the GLM t-statistics is shown in the two t-maps in Figure 7 (one from the GLM and one from the SVM created with the rescaled weight values). Finally, it is apparent that the strongest overall agreement between the two methods occur for lags 1 and 2 and that the biophysically implausible lag 0 shows significantly less impressive performance. Overall, this is a clear indication that the SVM is learning from the 15 participants, successfully predicting which experimental condition (tonal or atonal) is active at a given time point, and doing so on the basis of voxels which we know from the GLM analysis are significantly involved in the processing of those experimental conditions.

4. Discussion

We conducted an fMRI experiment on tonality in music and used this as an opportunity to explore the relationship between conventional GLM analysis and an SVM machine-learning approach. The detailed neuroanatomical results from a conventional GLM analysis alone (on all of the experiment conditions, rather than just the tonal–atonal contrast examined here), and an interpretation of these

results in the context of the existing literature, is not our purpose here and is reported elsewhere (Durrant et al. 2008). Our research here suggested that a successful SVM learning process required a considerable reduction in the very large dimensionality of the data set. As such, we used a group-level GLM analysis to first identify active voxels from a set of training data (15 of the participants), and then attempted to predict the prevailing stimulus conditions on the remaining subject from the BOLD signal alone. We also performed a GLM analysis on the remaining subject to identify active voxels relevant for the SVM task and obtained t-maps for this for each subject. Besides getting a measure of prediction performance from the SVM, therefore, we were also able to analyse at the individual voxel level the extent to which the two methods agreed upon which voxels were significant to distinguish between the experimental conditions and to what extent. Our results suggest that SVMs are capable of learning and generalising from fMRI data and predicting stimulus conditions, even when the data are very noisy (as is almost always the case for fMRI data), and the experimental task involves higher-order cognitive processing (of tonality, in the case of our experiment) rather than simple psychoacoustic tasks. This result is very encouraging for the use of machine learning in the automatic analysis of fMRI data, and the automatic Bayesian inference provided by the SVM highlights possible applications in the fledgling field of fMRI biofeedback among others. Many further developments of this technique are possible, in fine-tuning the algorithm and a more extensive search of the parameter space, in the possible use of GLM t-maps for prediction or SVM weights for regression and in an extension of the prediction to more than two conditions (involving multiple decision hyperplanes within the same weight space). The success of the SVM on this task, both in prediction and identification of relevant voxels, encourages further research in these areas.

Acknowledgements

This work was supported by the EPSRC project 'Learning the Structure of Music' (LeStruM),[3] EP-D063612-1. Also we would like to thank Janaina Mourão-Miranda for helpful advice. This publication only reflects the authors' views.

Notes

1. 'Atonal' here is used in the sense of 'not tonal', or 'having no key'. This should not be confused with the usage of 'atonal' in historical musicology, where it denotes music in a specific style and from a specific period.
2. The Spider SVM toolbox for MATLAB was used to perform the classifications: http://www.kyb.tuebingen.mpg.de/bs/people/spider/.
3. LeStruM project website http://www.lestrum.org/.

References

Baumgart, F., Kautisch, T., Tempelmann, C., Gaschler-Markefski, B., Tegeler, C., Schindler, F., Stiller, D., and Scheich, H. (1998), 'Electrodynamic Headphones and Woofers for Application in Magnetic Resonance Imaging Scanners', *Medical Physics* 25, 2068–2070.

Boser, B.E., Guyon, I.M., and Vapnik, V.N. (1992), 'A training algorithm for optimal margin classifiers', in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ACM*, pp. 144–152.

Carlson, T.A., Schrater, P., and He, S. (2003), 'Patterns of Activity in the Categorical Representations of Objects', *Journal of Cognitive Neuroscience*, 15, 704–717.

Cox, D.D., and Savoy, R.L. (2003), 'Functional Magnetic Resonance Imaging (fMRI) "Brain Reading": Detecting and Classifying Distributed Patterns of fMRI Activity in Human Visual Cortex', *Neuroimage*, 19, 261–270.

Cristianini, N., and Shawe-Taylor, J. (2000), *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge, UK: Cambridge University Press.

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D.G., Acharyya, M., Loughead, J.W., Gur, R.C., and Langleben, D.D. (2005), 'Classifying Spatial Patterns of Brain Activity with Machine Learning Methods: Application to Lie Detection', *NeuroImage*, 28, 663–668.

Durrant, S., Miranda, E., Brechmann, A., and Scheich, H. (2008), 'An fMRI Study of Neural Correlates of Musical Tonality and Atonality', *NeuroImage*, submitted for publication.

Eerola, T., and Toiviainen, P. (2004), *MIDI Toolbox: MATLAB Tools for Music Research*, Kopijyv, Jyvskyl, Finland: University of Jyvskyl. http://www.jyu.fi/musica/miditoolbox/.

Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., and Penny, W.D. (2006), *Statistical Parametric Mapping*, London: Academic Press.

Friston, K.J., Holmes, A.P., and Price, C.J. (1999), 'Multisubject fMRI Studies and Conjunction Analyses', *NeuroImage*, 19, 385–396.

Friston, K.J., Holmes, A.P., and Worsley, K.J. (1995), 'Statistical Parametric Maps in Functional Imaging: A General Linear Approach', *Human Brain Mapping*, 2, 189–210.

Haynes, J.D., and Rees, G. (2005), 'Predicting the Orientation of Invisible Stimuli from Activity in Human Primary Visual Cortex', *Nature Neuroscience*, 8, 686–691.

Janata, P., Birk, J.L., Van Horn, J.D., Leman, M., Tillmann, B., and Bharucha, J.J., (2002), 'The Cortical Topography of Tonal Structures Underlying Western Music', *Science*, 290, 2167–2170.

Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006), 'Information-based Functional Brain Mapping', *Proceedings of the National Academy of Sciences USA*, 103, 3863–3868.

Krumhansl, C.L. (1990), *Cognitive Foundations of Musical Pitch*, Oxford: OUP.

Krumhansl, C.L., and Kessler, E.J. (1982), 'Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys', *Psychological Review*, 89, 334–368.

LaConte, S., Strother, S., Cherkassky, V. Anderson, J., and Hu, X. (2005), 'Support Vector Machines for Temporal Classification of Block Design fMRI Data', *NeuroImage*, 26, 317–329.

Mitchell, T.M., Hutchinson, R., Niculescu, R.S., Pereira, F., Wang, X., Just, M., and Newman, S. (2004), 'Learning to Decode Cognitive States from Brain Images', *Machine Learning*, 1–2, 145–175.

Mourao-Miranda, J., Bokde, A.L.W., Born, C., Hampel, H., and Stetter, S. (2005), 'Classifying Brain States and Determining the Discriminating Activation Patterns: Support Vector Machine on Functional MRI Data', *NeuroImage*, 28, 980–995.

Narmour, E. (1991), *The Analysis and Cognition of Basic Melodic Structures: The Implication–Realization Model*, Chicago: University of Chicago Press.

Piston, W., and Devoto, M. (1987), *Harmony* (5th ed), New York: Norton.

Schmuckler, M.A., and Tomovski, R. (2005). 'Perceptual Tests of an Algorithm for Musical Key-finding', *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 1124–1149.

Shepard, R.N. (1982), 'Structural Representations of Musical Pitch', in *The Psychology of Music*, ed. D. Deutsch, Oxford, UK: OUP, pp. 344–390.

Tervaniemi, M. (2003), 'Musical Sound Processing: EEG and MEG Evidence', in *The Cognitive Neuroscience of Music*, eds. I. Peretz and R. Zatorre, New York: Oxford University Press, pp. 294–309.

Vapnik, V. (1995), *The Nature of Statistical Learning Theory*, New York: Springer-Verlag.

Wang, X., Hutchinson, R., and Mitchell, T.M. (2003),'Training fMRI Classifiers to Detect Cognitive States Across Multiple Human Subjects', in *Proceedings of the 2003 Conference on Neural Information Processing Systems*, Vancouver.