Interleaved silent steady state (ISSS) imaging: A new sparse imaging method applied to auditory fMRI

Christian Schwarzbauer,[a,b,*] Matt H. Davis,[a] Jennifer M. Rodd,[c] and Ingrid Johnsrude[a,d]

[a]MRC Cognition and Brain Sciences Unit, Cambridge, UK
[b]Department of Psychiatry, University of Cambridge, Cambridge, UK
[c]Department of Psychology, University College London, London, UK
[d]Department of Psychology, Queen's University, Kingston, Canada

The acoustic scanner noise that is generated by rapid gradient switching in echo planar imaging (EPI) is an important confounding factor in auditory fMRI. "Sparse imaging" designs overcome the influence of scanner noise on stimulus presentation by acquiring single brain volumes following a silent stimulus presentation period. However, conventional sparse imaging requires assumptions about the time-to-peak of the evoked hemodynamic response and reduces the amount of EPI data which can be acquired and hence statistical power. In this article, we describe an "interleaved silent steady state" (ISSS) sampling scheme in which we rapidly acquire a set of EPI volumes following each silent stimulus presentation period. We avoid T_1-related signal decay during the acquisition of the EPI volumes by maintaining the steady state longitudinal magnetization with a train of silent slice-selective excitation pulses during the silent period, ensuring that signal contrast is constant across successive scans. A validation study comparing ISSS to conventional sparse imaging demonstrates that ISSS imaging provides time course information that is absent in conventional sparse imaging data. The ISSS sequence has a temporal resolution like event-related (ER) imaging *within* a single trial (unlike conventional sparse imaging, where ER-like temporal resolution can only be achieved by compiling data across many jittered trials of the same stimulus type). This temporal resolution within trials makes ISSS particularly suitable for experiments in which a) scanner noise would interfere with the perception and processing of the stimulus; b) stimuli are several seconds in duration, and activation is expected to evolve and change as the stimulus unfolds; and c) it is impractical to present a single stimulus more than once (for example, repetition priming or familiarity effects would be expected).
© 2005 Elsevier Inc. All rights reserved.

Keywords: Auditory fMRI; Sparse imaging; Event-related fMRI; Acoustic noise

* Corresponding author. MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, CB2 2EF, UK.
E-mail address: christian.schwarzbauer@mrc-cbu.cam.ac.uk (C. Schwarzbauer).

Available online on ScienceDirect (www.sciencedirect.com).

Introduction

The acoustic scanner noise that is generated by rapid gradient switching in echo planar imaging (EPI) is an important confounding factor in auditory fMRI. The acoustic noise generated by a typical high-field MRI scanner is often sufficiently loud (in excess of 120 dB SPL) to make it difficult for subjects to hear auditory stimuli that might otherwise be expected to produce changes in BOLD activation. Even if stimuli can be detected over the acoustic noise of the scanner, it has been shown that the increased perceptual difficulty created by such background noise can substantially alter the pattern of activation obtained in simple contrasts (see Davis and Johnsrude, 2003 for an example in the perception of speech).

While a number of authors have proposed techniques by which the acquisition of fMRI data can be made quieter, all existing schemes introduce compromises either to the speed of acquisition, or the extent of coverage, or potentially to both (Loenneker et al., 2001; Amaro et al., 2002; Moelker and Pattynama, 2003). Although some authors have proposed the use of active noise-cancellation schemes to reduce the impact of scanner noise on stimulus perception, to date, there are no commercial noise-suppression systems that are MR-compatible nor will these systems ever completely compensate for bone-conducted (rather than air-conducted) scanner noise.

One widely used technique used to overcome the influence of scanner noise on stimulus presentation during fMRI temporally separates EPI scanner noise from the experimental sounds, taking advantage of the fact that the peak of the hemodynamic response lags the stimulus by several seconds (Edmister et al., 1999; Hall et al., 1999; Moelker and Pattynama, 2003). Data acquisition follows a silent interval during which experimental sounds are presented. Sequence timing is chosen such that the peak of the stimulus-related BOLD response coincides with data acquisition. However, this may be complicated by the variability of the time course of the BOLD response in different brain regions (Josephs and Henson, 1999). Furthermore, this experimental scheme is inefficient for

stimuli that vary in their length (such as words and sentences) as such stimuli may produce a peak hemodynamic response at different times. Finally, and related to this last point, conventional sparse imaging provides only a single measure of the hemodynamic response to a single stimulus and therefore does not permit differences in the timing or profile of the hemodynamic response to be assessed without scanning each stimulus multiple times.

In this article, we describe a sampling scheme that overcomes many of the limitations of the sparse imaging approach. A rapid set of n EPI volumes is acquired following each silent period. By acquiring multiple volumes, this protocol avoids the difficulties inherent in choosing, 'a priori', the time at which the evoked hemodynamic response will reach its peak. It also permits a greater proportion of scanning time to be used for data acquisition and an assessment of the time course of the hemodynamic response. To avoid T_1-related signal decay during data acquisition, the longitudinal magnetization is kept in a steady state by applying silent slice-selective excitation pulses during the silent period.

Methods

ISSS imaging

The principle of the ISSS imaging sequence is illustrated in Fig. 1. Each acquisition block consisted of a silent period during which the acoustic stimulus was presented and a data acquisition period. During the silent period, the longitudinal magnetization was kept in a steady state by applying a train of silent slice-selective excitation pulses using the same repetition time (TR) as during the data acquisition period. Silent slice-selective excitation was achieved by using a slice-selection gradient consisting of sinusoidal ramps and a constant plateau value during which the RF excitation pulse was applied (see Fig. 2). The gradient waveform was calculated according to

$$g(t) = G_S \begin{cases} 0.5 - 0.5\cos\left(\frac{\pi t}{T_{Ramp}}\right) & 0 < t < T_{Ramp} \\ 1 & T_{Ramp} < t < T_{Ramp} + T_P \\ 0.5 + 0.5\cos\left(\frac{\pi(t - T_{Ramp} - T_P)}{T_{Ramp}}\right) & T_{Ramp} + T_P < t < 2T_{Ramp} + T_P \end{cases} \quad (1)$$

where T_{Ramp} and T_P are the corresponding durations of the gradient ramp and plateau, and G_S is the plateau value of the slice-selection gradient. Gradient waveforms of this kind have
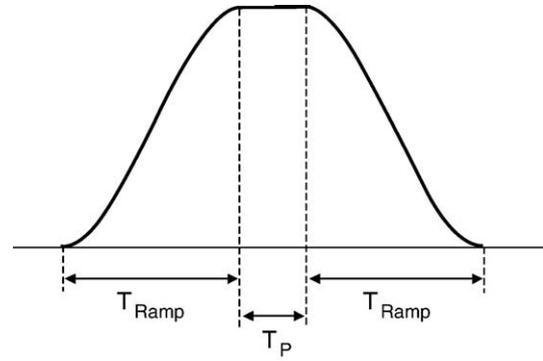

Fig. 2. The silent slice-selection gradient as defined by Eq. (1). It consists of sinusoidal ramps (T_{Ramp} = 3 ms) and a constant plateau value (T_P = 1 ms).

been demonstrated to produce a maximally band-limited acoustic frequency response, which essentially contains only frequencies below

$$f_{max} = 2/T_{Ramp}, \quad (2)$$

the power of harmonics above this cut-off limit being attenuated by more than 30 dB (Hennel et al., 1999). As the acoustic frequency response function of typical gradient coils is very low (in the range below 200 Hz), acoustic noise produced by the gradient switching will be efficiently attenuated in this frequency range (efficient acoustic attenuation range = EAAR). A further advantage is that the sensitivity of the human hearing in this spectral range is considerably reduced. Efficient acoustic noise reduction can therefore be achieved by choosing T_{Ramp} such that f_{max} lies within the EAAR of the gradient coil. If the acoustic frequency response function of the gradient system is known, the minimum T_{Ramp} value can, in principle, be derived theoretically. In the present study, however, we took a more pragmatic approach that does not require any special equipment. Starting at 1.0 ms, T_{Ramp} was increased in steps of 0.5 ms until the clicking noise produced by the switching of the slice gradient was indistinguishable from the background noise. This was the case at T_{Ramp} = 3.0 ms. The duration of the gradient plateau was T_P = 1.0 ms. The total time required for the silent slice-selective excitation therefore amounted to $2T_{Ramp} + T_P$ = 7 ms. We used the same silent slice-selective excitation scheme for both dummy and true scans. For true scans, an additional refocusing gradient was applied to remove the phase dispersion generated as a result of spin dephasing during slice-selective excitation. In principle, silent slice-selective excitation is unnecessary for true scans as the
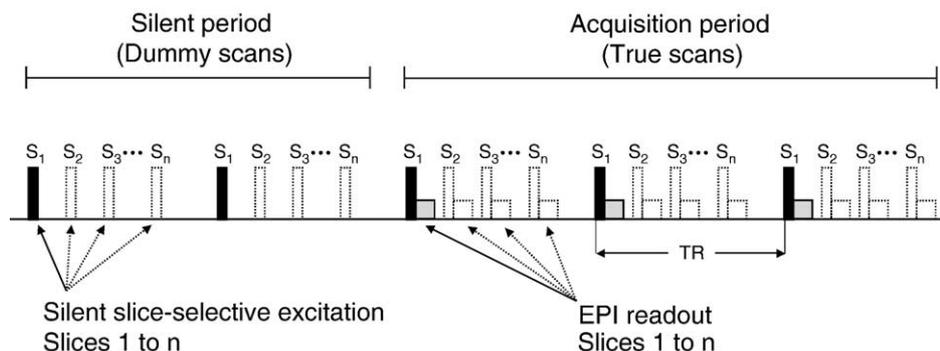

Fig. 1. Schematic representation of the multi-slice ISSS imaging sequence. Silent periods, which are an integer multiple of the TR, are interleaved with periods of data acquisition. During the silent periods, the steady state longitudinal magnetization is maintained with silent slice-selective excitation pulses.

EPI readout produces a considerable amount of acoustic noise anyway. However, using the same excitation scheme for both dummy and true scans rules out instabilities in the steady state signal due to a mismatch in the corresponding slice profiles.

Materials

Conventional sparse imaging is used to detect relatively stable increases in BOLD signal resulting from auditory stimuli that are extended in time. One goal of this validation study was to compare the sensitivity of conventional sparse imaging with that of the new ISSS sequence to such 'steady state' increases in BOLD. This involved comparing two auditory conditions: listening to simple spoken sentences and listening to modulated noise with similar acoustic properties. Materials for this study were a set of 60 sentences taken from an earlier imaging study (Rodd et al., 2005; low-ambiguity sentences). These were between 7 and 13 words in length and ranged in duration from 1.7 to 4.3 s. An equal number of length-matched sentences was also converted into 'signal-correlated noise' (SCN; Schroeder, 1968) using Praat software (www.praat.org). SCN sentences retain the amplitude envelope and spectral profile of the original speech but are completely unintelligible and therefore provide an appropriate baseline for detecting sentence-specific activation (Mummery et al., 1999; Davis and Johnsrude, 2003; Rodd et al., 2005). Sixty "silent" trials were also scanned to permit comparison of activation during the auditory conditions with that during a resting baseline.

The second goal was to establish that the ISSS sequence would be sensitive to a phasic or transient response. This involved contrasting an auditorily cued button press with rest. Thus, on 40 of the 60 trials in each of the three conditions (sentences, noise and silence), a short tone (sinusoidal, 880 Hz, 200 ms duration) cued a button press with the index finger of the right hand. The tone sounded either 5 s (early) or 2 s (late) before the onset of the subsequent true scans for the ISSS sequence and 7 s (early) or 4 s (late) before the onset of the single scan for the conventional sparse sequence (see Fig. 3). One third of the trials in each condition did not contain a tone, providing matched baseline conditions against which to compare early and late button press trials. The 180 trials were presented in pseudorandom order divided into four runs of 45 trials each.

Auditory stimuli were presented using Resonance Technology Commander XG 'Ultra-Fidelity' headphones incorporating some sound attenuation. Gradient EPI on the Bruker scanner is exceedingly noisy; scanner noise peaks at levels in excess of 120 dB. Therefore, to further attenuate scanner noise, volunteers also wore insert earplugs (E.A.R. Supersoft; Aearo Company, Indianapolis, IN) rated to attenuate by ~30 dB. When wearing ear plugs and the ear defenders of the Commander XG system, participants reported that the scanner noise (during acquisition) was unobtrusive and that sentence and noise stimuli were presented at a comfortable listening volume. During silent periods of the ISSS sequence, there was no additional noise from the scanner over and above the noise of the cooling system pumps that are continuously in operation.

Participants

Data were acquired from 5 right-handed, female, neurologically normal participants, all of whom were native speakers of British English with normal or corrected-to-normal vision, all reported having normal hearing and were not observed to suffer from any language deficit. This study was approved by the Addenbrooke's Local Research Ethics Committee (Cambridge, UK), and written informed consent was obtained from all participants. Volunteers were told that they would be hearing sentences or bursts of noise on most trials, but they should listen out for a short tone which would be present on two thirds of trials and press a button with their right hand whenever they heard it.

Magnetic resonance imaging

Imaging was performed on a 3 T Bruker Medspec system (Bruker Biospin GmbH, Ettlingen, Germany), using a head gradient insert and a quadrature birdcage head coil for transmit and receive.

Each participant was scanned for two runs with each of two sequences (ISSS and conventional sparse imaging; see Fig. 3) in counterbalanced order (except for one participant who was only scanned for a single run using each sequence). Each scanning run
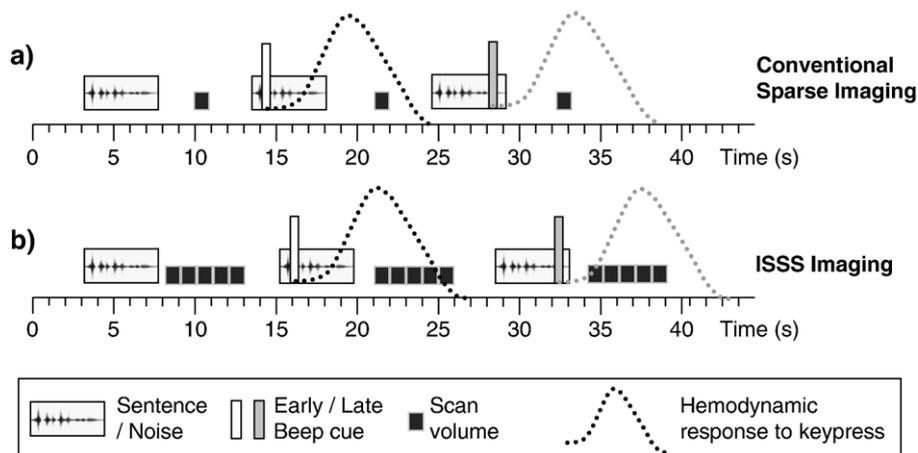

Fig. 3. Timelines of the imaging protocols used in this study: (a) conventional sparse imaging; (b) ISSS imaging. By acquiring multiple volumes, this ISSS scheme avoids the difficulties inherent in choosing, 'a priori', the time at which the evoked hemodynamic response will reach its peak. It also permits a greater proportion of scanning time to be used for data acquisition and assessment of the time course of the hemodynamic response. During the silent periods, the longitudinal magnetization is kept in a steady state to avoid T_1-related saturation effects.

consisted of the 45 of the stimuli described above: participants listened to sentences, modulated noise or silence in order to detect a short tone that cued a button press. The conventional sparse imaging sequence consisted of a 1-s scan acquired 2 s after the offset of each spoken sentence. The scan was repeated 45 times (TR = 11 s) resulting in a total acquisition time of 8 min and 15 s (45 volumes were acquired). The new ISSS sequence consisted of 5 1-s scans that were acquired at a repetition time of TR = 1 s, starting immediately after the offset of each stimulus. During the presentation of the auditory stimuli, 8 dummy scans were performed (TR = 1 s) to keep the longitudinal magnetization in a steady state. Each dummy scan consisted of a silent slice-selective excitation pulse as described above. The duration of this acquisition block was T_{ISSS} = 13 s, and it was repeated 45 times, giving a total acquisition time of 9 min and 45 s (225 volumes were acquired). This design allows us to compare equivalent data taken using the two sequences since the third acquisition in each block of ISSS scans had the same timing relative to stimulus presentation and response measurement as the single scan acquired using the conventional sparse imaging design. The timeline of these events for both conventional sparse imaging and the ISSS sequence is depicted in Fig. 3.

For both imaging sequences, the following imaging parameters were used: slice thickness, 4 mm; inter-slice gap, 1 mm; number of slices, 18; slice orientation, axial oblique; field of view, 20 × 20 cm, matrix size, 64 × 64, in-plane spatial resolution 3.1 × 3.1 mm; acquisition bandwidth, 101 kHz; and echo time, 27.5 ms. The optimum excitation slice angles (Ernst angle) were calculated according Eq. (5) to give α = 63° (ISSS sequence, TR = 1 s) and α = 90° (conventional sparse imaging sequence, TR = 11 s), assuming a constant T_1 value of 1.25 s for gray matter at a field strength of 3 T (Norris, 2000). For the ISSS sequence, the gradient ramp and plateau times were T_{Ramp} = 3.0 ms and T_P = 1.0 ms (see ISSS imaging).

Subsequent to functional acquisitions, field map data to facilitate geometric undistortion during preprocessing were acquired (Cusack and Papadakis, 2002). For spatial normalization, structural images were acquired using a 3D T_1-weighted SPGR sequence with 1 mm isotropic spatial resolution.

Considerations regarding the signal-to-noise ratio

In ISSS imaging, the longitudinal magnetization is kept in a steady state for the duration of the entire experiment. Therefore, only a fraction of the equilibrium magnetization will be available for image formation resulting in a reduced signal-to-noise ratio (SNR) per image volume. In terms of spin dynamics, both the ISSS and the conventional sparse imaging sequence can be considered to be spoiled gradient-echo sequences. The relative SNR (i.e. the SNR with respect to the equilibrium state) is therefore given by (Haacke et al., 1999)

$$SNR/SNR_0 = \frac{1 - E_1}{1 - E_1 \cos(\alpha)} \sin(\alpha) \tag{3}$$

with

$$E_1 = \exp(-TR/T_1) \tag{4}$$

It is straightforward to show that the maximum SNR/SNR_0 value is obtained if

$$\cos(\alpha) = E_1 \tag{5}$$

The resulting optimum flip angle is also known as the Ernst angle (Haacke et al., 1999). Using this expression, Eq. (3) can be rewritten to give

$$SNR/SNR_0 = \sqrt{(1 - E_1)/(1 + E_1)} \tag{6}$$

Assuming a constant T_1 value of 1.25 s for gray matter at a field strength of 3 T (Norris, 2000), SNR/SNR_0 can be calculated from Eqs. (4) and (6). For the ISSS imaging sequence (TR = 1 s), we obtain SNR/SNR_0 = 0.62. For the conventional sparse imaging sequence (TR = 11 s), the corresponding values is SNR/SNR_0 = 1.0. In the latter case, the full equilibrium magnetization is available for image formation as a result of the long TR (see Eqs. (4) and (6)).

So far, we have only considered the SNR per image volume. For a direct comparison of both sequences, however, it is more appropriate to consider the time-normalized SNR given by $SNR/\sqrt{T}$ (Haacke et al., 1999), where T denotes the time needed to acquire a certain amount of information. Typically, this would be an image slice or volume. If, for example, sequence A and B provided the same SNR per image volume, but sequence A was 4 times faster, the corresponding $SNR/\sqrt{T}$ would be twice the one obtained with sequence B. The ratio

$$\varepsilon = \frac{SNR_A/\sqrt{T_A}}{SNR_B/\sqrt{T_B}} \tag{7}$$

therefore provides a useful measure to compare the efficiency of two imaging sequences in terms of their relative $SNR/\sqrt{T}$. With the ISSS sequence, n = 5 points of the hemodynamic response are sampled following the presentation of a stimulus. The time needed to acquire this information is T_{ISSS} = 13 s. In conventional sparse imaging, only a single point of the hemodynamic response can be sampled following the presentation of a stimulus. In order to obtain the same temporal resolution as obtained with the ISSS sequence, the conventional sparse imaging sequence would have to be repeated n = 5 times using similar stimuli but different times between stimulus presentation and volume acquisition to sample a different point of the hemodynamic response each time (Belin et al., 1999). This principle is often referred to as 'jittering'. The repetition time of such a 'jittered' sequence is determined by the longest time interval between stimulus presentation and volume acquisition, which exactly corresponds to the duration of one ISSS acquisition. The time needed to sample n = 5 points of the hemodynamic response would therefore be nT_{ISSS} = 65 s. With $T_A = T_{ISSS}$ and $T_B = nT_{ISSS}$, the efficiency of the ISSS sequence (with respect to the jittered sparse imaging sequence) can be calculated according to Eq. (7) to give

$$\varepsilon = \sqrt{n} \frac{SNR_{ISSS}}{SNR_J} \tag{8}$$

For n = 5, SNR_{ISSS} = 0.62 SNR_0 and SNR_J = 1.0 SNR_0, we obtain ε = 1.39. This means that, for the specific choice of imaging parameters used in this study, the ISSS sampling scheme is 39% more efficient than a jittered sparse imaging scheme.

For a more general comparison of both sequences, ε was calculated as a function of n for 3 different temporal resolutions (Δt = 0.7, 1.0 and 1.5 s). Calculations were performed according to Eqs. (4), (6) and (8), using TR = 13 s (jittered sparse imaging sequence); TR = Δt (ISSS sequence); and T_1 = 1.25 s (see paragraph following Eq. (6)). The results are illustrated in Fig. 6 and will be discussed at a later stage.

Analysis

Data from both acquisition protocols were preprocessed and analyzed using SPM2 (http://www.fil.ion.ucl.ac.uk/spm/). Preprocessing stages included realignment and unwarping to correct for subject motion and interactions between movement and field inhomogeneities (Andersson et al., 2001). The unwrapped phase and magnitude of the field maps were then spatially processed and scaled (Cusack et al., 2003) and used to remove geometric distortions from the EPI images. Images were then coregistered to a structural image and normalized to the MNI template brain followed by smoothing at 10 mm FWHM.

Images were entered into single-subject GLMs using SPM2 software. Data from each scanning run for sparse imaging were modeled using a design matrix consisting of five columns which indicated whether scans were preceded by presentation of a (1) sentence or (2) signal-correlated noise, as well as scans preceded by (3) early or (4) late cues for button presses or (5) error trials on which a cue was missed or a false alarm was made. Silent baseline scans and scans not preceded by a button press are therefore unmodeled baseline scans in this analysis. Regressors of no interest encoded scan-by-scan movement parameters from the realignment stage of preprocessing and a constant term for each scanning session. The same events and regressors were modeled for data collected using the ISSS sequence; however, each of the 5 event types was modeled over 5 separate columns, a single column for each of the 5 scans acquired after each silent period.

Contrasts of parameter estimates for the least-squares fit of these models in each subject were then taken forward into second-level random-effects analyses (one-sample t tests) in which session was treated as a random variable. By comparing activation over sessions in this way, we avoid the thorny issue of whether the multiple scans collected after each silent period in the ISSS sequence are statistically independent. In all cases, these random-effects analyses are assessed with 8 degrees of freedom (since we collected and analyzed 9 scanning runs from each sequence).

Results

All subjects performed the behavioral task correctly, with two out of five participants making no errors, a further two participants omitting one button press and a fifth making three false alarms. Error trials were modeled separately in those participants who made them, so results reflect only correct performance of the button press task.

Regardless of sequence used to acquire the data, contrasts between speech and signal-correlated noise revealed activation bilaterally in the superior temporal gyrus and superior temporal sulcus (more extensive on the left), consistent with earlier findings (Davis and Johnsrude, 2003; Rodd et al., 2005). Activation foci at a whole-brain-corrected level of significance were observed in both anterior and posterior temporal regions (see Figs. 4a, b and Table 1) for both ISSS and conventional sparse imaging sequence (Figs. 4a, b). An analysis that directly compared activation for speech over signal-correlated noise in the conventional and ISSS sequence showed no significant differences in activation in the two sequences at even a liberal statistical threshold (no voxels at $P < 0.05$ uncorrected).

Comparisons of plots of the peak response in the anterior superior temporal gyrus for both the conventional and ISSS
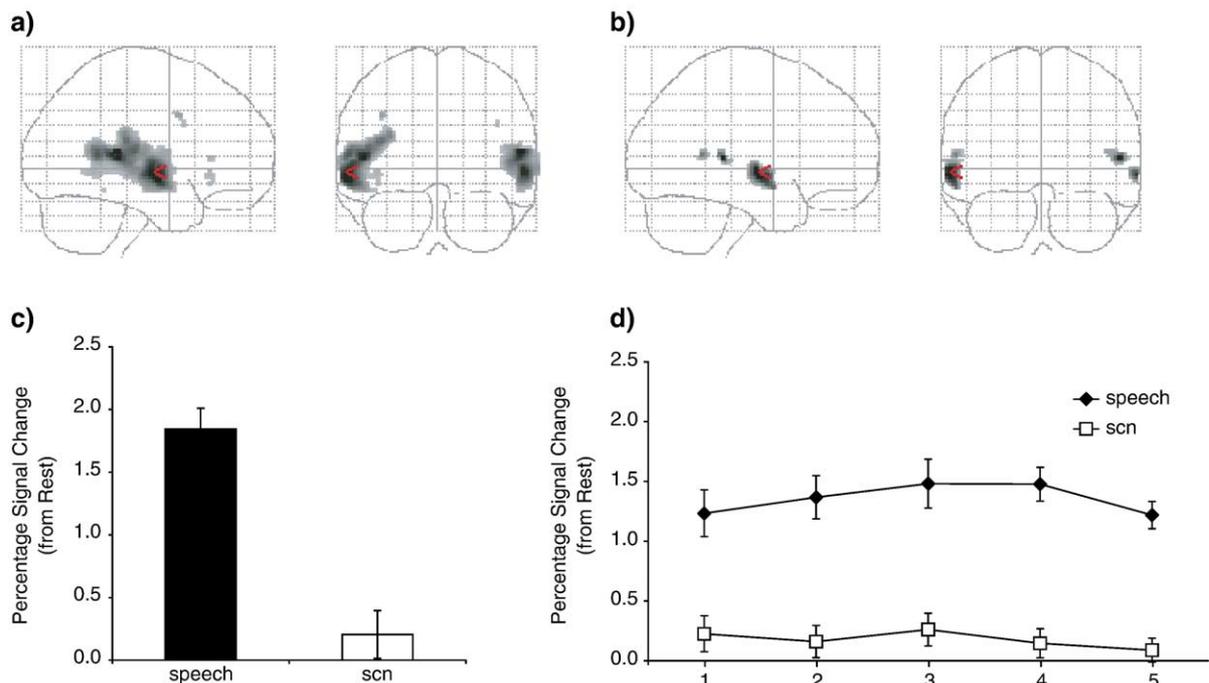

Fig. 4. (a) Glass-brain depiction of random-effects analyses ($n = 9$) contrasting scans following speech and signal-correlated noise (SCN), thresholded at $P < 0.05$ FDR, for conventional sparse imaging; (b) as before for ISSS imaging. (c) Plot of mean percent signal change (compared to rest) in scans following speech and signal-correlated noise for a peak voxel in the left anterior superior temporal sulcus ($x = -62$; $y = -10$; $z = -4$) for conventional sparse imaging. Error bars show the standard error of the mean over conditions. (d) Signal change plotted as before for ISSS imaging ($x = -62$; $y = -10$; $z = -4$). Each data point represents the response in 1 of the 5 scans following each silent interval.

Table 1
Contrast between speech and signal-correlated noise for ISSS and conventional sparse imaging

Brain area	L/R	ISSS sparse				Conventional sparse			
		x	y	z	Z	x	y	z	Z
Superior temporal gyrus (anterior)	L	**−62**	**−10**	**−4**	**4.91**	**−62**	**−10**	**−4**	**4.79**
	R	+64	−4	−4	4.70	−58	−6	−2	4.71
Superior temporal gyrus (posterior)	L	−58	−48	+10	4.31	−60	−48	+10	4.43
		−64	−30	+6	3.98	−50	−34	+8	4.81
	R	+52	−34	+8	4.80	+56	−34	+12	4.91
		+44	−34	+12	3.99	+60	−24	0	3.93
						+66	−48	+10	3.11
Inferior frontal gyrus (partes opercularis and orbitalis)	L					−38	+8	+36	3.56
						−42	+28	−4	3.21
	R					+56	+28	+8	3.17
						+40	+10	+30	3.06
Inferior temporal gyrus (posterior)	L					−44	−44	−14	3.07

Table reports peak voxels (more than 8 mm apart) that exceed a whole-brain, FDR-corrected threshold of $P < 0.05$ (Genovese et al., 2002). The response of peak voxels shown in bold type is plotted in Figs. 4c and d.

sequences (shown in Figs. 4c and d) confirm that the significant activation observed for speech compared to signal-correlated noise in the single sparse imaging scan reflects sustained activity over a number of scans in the ISSS sequence. Comparing the response in the sparse imaging scan with that in the equivalent single scan for the ISSS sequence (scan 3) demonstrates a reduction of approximately 30% in the magnitude of the local signal change for speech or SCN compared to rest (see Figs. 4c, d).

We now turn to the contrast between cued button press responses and matched non-cue trials in which participants listened passively to the same stimuli. This allows us to assess the ability of the ISSS sequence to detect transient activity associated with the cued button press. We conducted a region-of-interest analysis to assess activation in the left motor cortex (corresponding to the right-hand button presses, either 5 or 2 s before the onset of the ISSS true scans, that participants executed). The region of interest defined for this analysis was based on the mean location of the left 'hand-knob' landmark for motor cortex in a group of normal volunteers (Yousry et al., 1997). We defined a spherical ROI with a radius of 12 mm around the reported coordinates (centered on $x = -36$, $y = -18$, $z = +54$). Within this region of interest, we observed a number of voxels showing an elevated response to trials with a button press response compared to those without, significant at an FDR small-volume-corrected level (Worsley et al., 1996; reported in Table 2).

Plots of signal change in the peak motor cortex voxel for the two sequences (Figs. 5a and b) show that the conventional sparse

Table 2
Contrast between button press and non-button press trials for ISSS and conventional sparse imaging sequences

Brain area	L/R	ISSS sparse				Conventional sparse			
		x	y	z	Z	x	y	z	Z
Motor cortex	L	−30	−26	+48	**3.51**	−34	−26	+46	**3.84**
		−38	−20	+56	3.40	−38	−24	+62	3.71
		−28	−18	+56	3.07	−32	−10	+62	2.71

No voxels exceed a whole-brain, FDR-corrected threshold of $P < 0.05$ (Genovese et al., 2002). Peak voxels are reported (more than 8 mm apart) that reach an FDR-corrected level of significance within a 12 mm spherical ROI centered on the hand region of left motor cortex ($x = -36$, $y = -18$, $z = +54$, from Yousry et al., 1997). Response of peak voxels shown in bold type is plotted in Figs. 5a and b.

imaging sequence gives the impression that late button presses produce a large BOLD response, whereas responses to early button presses are weak or non-existent. However, examination of the time course of motor cortex activity in the ISSS data reveals that motor cortical activity related to the button press is transient. The evoked response to early and late button presses is of approximately equal magnitude but differs in the timing of the peak response. The result in the conventional sparse data is therefore misleading and is an artefact of stimulus timing relative to scanning. The 'late' button press occurs ~4 s before the onset of the single conventional scan, and so the evoked hemodynamic response would be expected to be near maximum during scanning (since the canonical hemodynamic response function peaks at ~5 s). In contrast, the 'early' button press is occurring ~7 s before scanning begins, and the evoked hemodynamic response would be sampled after the peak.

Discussion

The results of this study demonstrate the potential of the ISSS sequence to track the time course of activation produced by acoustic stimuli over a period of several seconds. We have demonstrated that, with a rapid acquisition (TR = 1 s), we can clearly distinguish the temporal order of events separated by only 3 s. Such results are extremely promising for studies in which the goal is to measure the evolution of activation over time during auditory stimuli such as sentences, musical phrases or environmental sounds, which might be expected to recruit different cognitive processes at different points.

In ISSS imaging, the longitudinal magnetization is kept in a steady state in order to avoid T_1-related signal decay during the acquisition period. This is achieved by applying a train of silent slice-selective excitation pulses during the silent period using the same repetition time as during the acquisition period. In the absence of the slice-selective excitation pulses during the silent period, the signal would decrease during the acquisition period due to the progressive saturation of the longitudinal magnetization. Depending on the repetition time, this effect can be substantial. For the acquisition parameters used in this study, both the normalized signal intensity and SNR would drop from 100% to 62% during the acquisition of the time course. A direct observation of
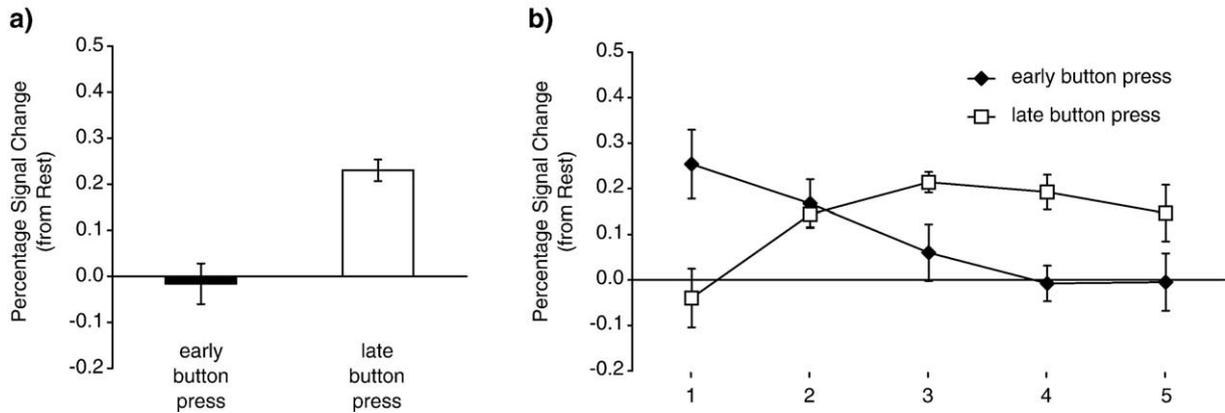
Fig. 5. (a) Plot of mean percent signal change (compared to rest) in scans following early and late button presses in the left motor cortex ($x = -34$; $y = -26$; $z = +46$) for conventional sparse imaging. Error bars show the standard error of the mean over conditions. (b) As before for ISSS imaging ($x = -30$; $y = -26$; $z = +48$). Each data point represents the response in 1 of the 5 scans following each silent interval.

stimulus-related signal changes, which are typically only of the order of a few percent, would therefore be very difficult. This problem has been addressed in some recent publications, and it has been demonstrated that functional activation can be detected if the initial signal drop is modeled appropriately (Bandettini et al., 1998; Guimaraes et al., 1998). The drawback of this approach, however, is the drop in the SNR which results in a different functional sensitivity for each point of the time course of hemodynamic response. Conversely, the ISSS sequence provides the same SNR and therefore the same functional sensitivity for all data points.

In conventional sparse imaging, only a single image volume is acquired after the presentation of a stimulus. In order to obtain information about the time course of the hemodynamic response, it has been proposed to jitter the timing of the presentation of the auditory stimulus relative to the clustered volume acquisition (Belin et al., 1999). However, there are several disadvantages associated with this approach. First, if the time course of the hemodynamic response is to be sampled at n different points, the presentation of each stimulus needs to be repeated n times during the experiment which may produce a reduced hemodynamic response due to repetition suppression (see Henson, 2003 for a review). Any change in the hemodynamic response due to repetition challenges a necessary assumption of the jittering approach—that the hemodynamic response for a given stimulus remains the same no matter how often this stimulus is repeated during the experiment. No such assumption is required for the ISSS sequence since the hemodynamic response can be sampled at multiple time points after a single event. Second, in terms of the time-normalized SNR, the 'jittered' sparse imaging approach is generally less efficient than the ISSS sampling scheme as illustrated in Fig. 6. For the imaging parameters we used in our study, for example, the ISSS sequence is 1.39 times more efficient than the jittered sequence (see filled circle in Fig. 6). In general, the efficiency (ε) of the ISSS sequence increases with the number of data points (n) sampled. Only if very few data points are sampled does the ISSS sequence becomes less efficient than the jittered sequence ($\varepsilon < 1$). For TR = 1 s, for example, this is only the case if $n \leq 2$. This property reflects a trade-off between the benefits of acquiring more data points in the ISSS sequence and the reduced SNR of each ISSS volume (see Considerations regarding the signal-to-noise ratio). What is more, a trade-off of this kind would

even be possible at a post-processing stage. The SNR could be simply improved by a factor of $\sqrt{k}$ by averaging k subsequent image volumes, but this would come at the cost of reducing the temporal resolution by a factor of k. Instead of averaging, digital filtering could be applied in the time domain. Thus, even non-integer k values would be possible.

Comparing the response in the sparse imaging scan with that in the equivalent single scan for the ISSS sequence (scan 3), we observed a reduction of approximately 30% in the fractional signal change for speech or SCN compared to rest (see Figs. 4c, d). For this particular region of interest, the ~30% decrease in fractional signal change would almost cancel out the benefits of the ISSS sequence in terms of its better efficiency. The observed fractional signal decrease may be due to the different T_1 weighting of both sequences (Lu et al., 2002). Interestingly, Lu et al. reported both signal increases and decreases with increased T_1 weighting. This effect could be described by an effective regional efficiency that varies around the theoretical value calculated according to Eq. (8).

One limitation of this ISSS technique concerns the duration of time for which we might expect activation to be uncontaminated by
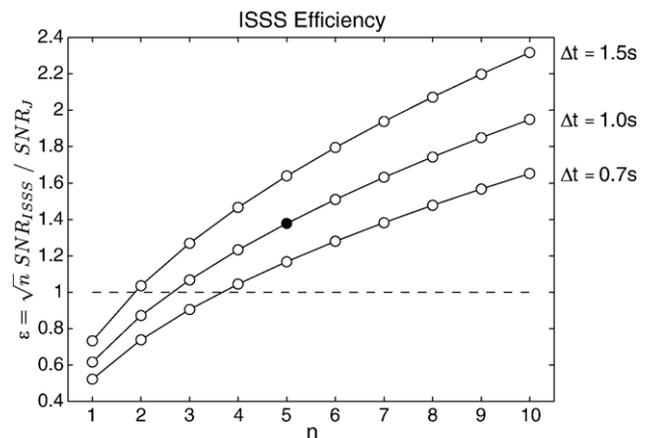

Fig. 6. The efficiency (ε) of the ISSS acquisition scheme with respect to a jittered sparse imaging scheme plotted as a function of the number of points (n) sampled of the time course of the hemodynamic response. Plots are shown for three different temporal resolutions (Δt). The filled circle indicates the efficiency obtained for the imaging parameters used in this study.

the neural response to scanner noise. The time course of the hemodynamic response (e.g. Hall et al., 1999) indicates that the neural response to scanner noise would start to become apparent in scans taken more than 4–5 s after the onset of scanner noise. Similarly, as a consequence of the hemodynamic delay, this sequence will have only a limited ability to detect neural events that occurred more than 5 to 7 s before the initial scan. However, these limitations fit very nicely within the time course of a number of auditory processes of interest. For instance, a typical spoken sentence might last 2 or 3 s—a duration which is well suited to imaging studies that use the ISSS sequence. The buildup of auditory streaming also has a similar time course (Carlyon et al., 2001; Cusack, 2005), and mismatch-negativity studies (Naatanen and Winkler, 1999) similarly require a few seconds of auditory 'standard' stimuli in order to produce an elevated response to an occasional deviant stimulus. Where researchers are interested in auditory processes that evolve over an even longer duration, it would be possible to tailor the duration of the dummy scan (silent) period and the number of acquisitions in the ISSS time-series to the stimuli and processes under study. In general, the durations of the silent and acquisition periods can be any integer multiple of the basic TR (in our case, 1 s; see Fig. 1). Furthermore, silent and acquisition periods of variable durations could be used during an experimental run to accommodate more complex experimental designs.

One issue that arises in any use of sparse imaging with stimuli longer than 2 or 3 s in duration is that there may be only limited opportunity to sample the initial onset of the hemodynamic response. This is true for all sparse imaging, whether multiple scans are acquired or not. The need to present stimuli during silence prevents any data being acquired during this period of the steepest rise in the hemodynamic response. However, while some researchers have sought to measure the latency of the BOLD response (Henson et al., 2002; Bellgowan et al., 2003), most current fMRI applications depend only on comparing the magnitude of stimulus-induced activation across conditions. By appropriately varying the timing of events during the silent pre-stimulus period, it would be possible to acquire the same data with a conventional sparse imaging sequence, although this would require an increase in the number of repetitions of each stimulus. By comparison, the ISSS sequence proposed here is therefore to be favored for situations in which stimulus presentation without simultaneous scanner noise is desired and neural correlates of events occurring at variable times during an auditory sequence are to be detected.

Conclusion

In summary, we have demonstrated a novel EPI acquisition protocol with considerable potential for use in auditory fMRI studies. The combination of an extended silent period, followed by rapid acquisition of multiple brain volumes, provides an optimal compromise between two conflicting requirements: (1) the need to present auditory stimuli against a silent background and (2) the need to collect multiple rapid acquisitions in order to efficiently and accurately measure the hemodynamic response. Our results demonstrate the suitability of the ISSS sequence for detecting long-lasting auditory and temporal cortical activation evoked by sentence presentation and the more transient motor cortical activation evoked by auditorily cued button presses. In the latter case, ISSS imaging provided time course information that was absent from traditional sparse imaging procedures. ISSS imaging is a practical and useful tool for both scientific and clinical applications. In a scientific context, collecting more data per unit time is helpful. In a clinical context, reducing scan time may be more important. In both cases, ISSS offers significant advantages.

References

Amaro, E., Williams, S.C.R., Shergill, S.S., Fu, C.H.Y., MacSweeney, M., Picchioni, M.M., Brammer, M.J., McGuire, P.K., 2002. Acoustic noise and functional magnetic resonance imaging: current strategies and future prospects. J. Magn. Reson. Imaging 16, 497–510.

Andersson, J.L., Hutton, C., Ashburner, J., Turner, R., Friston, K., 2001. Modeling geometric deformations in EPI time series. NeuroImage 13, 903–919.

Bandettini, P.A., Jesmanowicz, A., Van Kylen, J., Birn, R.M., Hyde, J.S., 1998. Functional MRI of brain activation induced by scanner acoustic noise. Magn. Reson. Med. 39 (3), 410–416.

Belin, P., Zatorre, R.J., Hoge, R., Evans, A.C., Pike, B., 1999. Event-related fMRI of the auditory cortex. NeuroImage 10, 417–429.

Bellgowan, P.S.F, Saad, Z.S., Bandettini, P.A., 2003. Understanding neural system dynamics through task modulation and measurement of functional MRI amplitude, latency, and width. Proc. Natl. Acad. Sci. U. S. A. 100 (3), 1415–1419.

Carlyon, R.P., Cusack, R., Foxton, J.M., Robertson, I.H., 2001. Effects of attention and unilateral neglect on auditory stream segregation. J. Exp. Psychol. Hum. Percept. Perform. 27, 115–127.

Cusack, R., 2005. The intraparietal sulcus and perceptual organization. J. Cogn. Neurosci. 17, 641–651.

Cusack, R., Papadakis, N., 2002. New robust 3-D phase unwrapping algorithms: application to magnetic field mapping and undistorting echoplanar images. NeuroImage 16, 754–764.

Cusack, R., Brett, M., Osswald, K., 2003. An evaluation of the use of magnetic field maps to undistort echo-planar images. NeuroImage 18, 127–142.

Davis, M.H., Johnsrude, I.S., 2003. Hierarchical processing in spoken language comprehension. J. Neurosci. 23, 3423–3431.

Edmister, W.B., Talavage, T.M., Ledden, P.J., Weisskoff, R.M., 1999. Improved auditory cortex imaging using clustered volume acquisitions. Hum. Brain Mapp. 7, 89–97.

Genovese, C.R., Lazar, N.A., Nichols, T.E., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15, 870–878.

Guimaraes, A.R., Melcher, J.R., Talavage, T.M., Baker, J.R., Ledden, P., Rosen, B.R., Kiang, N.Y.S., Fullerton, B.C., Weisskoff, R.M., 1998. Imaging subcortical auditory activity in humans. Hum. Brain Mapp. 6, 33–41.

Haacke, E.M., Brown, R.W., Thompson, M.R., Venkatesan, R., 1999. Magnetic Resonance Imaging: Physical Principles and Sequence Design. Wiley, New York.

Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., Gurney, E.M., Bowtell, R.W., 1999. "Sparse" temporal sampling in auditory fMRI. Hum. Brain Mapp. 7, 213–223.

Hennel, F., Girard, F., Loenneker, T., 1999. "Silent" MRI with soft gradient pulses. Magn. Reson. Med. 42, 6–10.

Henson, R.N.A., 2003. Neuroimaging studies of priming. Prog. Neurobiol. 70, 53–81.

Henson, R.N.A., Price, C., Rugg, M.D., Turner, R., Friston, K., 2002. Detecting latency differences in event-related BOLD responses: application to words versus nonwords, and initial versus repeated face presentations. NeuroImage 15, 83–97.

Josephs, O., Henson, R.N.A., 1999. Event-related functional magnetic resonance imaging: modelling, inference and optimization. Philos. Trans. R. Soc. London, Ser. B Biol. Sci. 354, 1215–1228.

Loenneker, T., Hennel, F., Ludwig, U., Hennig, J., 2001. Silent BOLD imaging. Magn. Reson. Mater. Phys., Biol., Med. 13, 76–81.

Lu, H., Golay, X., van Zijl, C.M., 2002. Intervoxel heterogeneity of event-related functional magnetic resonance imaging responses as a function of T_1 weighting. NeuroImage 17, 943–955.

Moelker, A., Pattynama, P.M.T., 2003. Acoustic noise concerns in functional magnetic resonance imaging. Hum. Brain Mapp. 20, 123–141.

Mummery, C.J., Ashburner, J., Scott, S.K., Wise, R.J., 1999. Functional neuroimaging of speech perception in six normal and two aphasic subjects. J. Acoust. Soc. Am. 106, 449–457.

Naatanen, R., Winkler, I., 1999. The concept of auditory stimulus representation in cognitive neuroscience. Psychol. Bull. 125, 826–859.

Norris, D.G., 2000. Reduced power multislice MDEFT imaging. J. Magn. Reson. Imaging 11, 445–451.

Rodd, J.M., Davis, M.H., Johnsrude, I.S., 2005. The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. Cereb. Cortex 15, 1261–1269.

Schroeder, M.R., 1968. Reference signal for signal quality studies. J. Acoust. Soc. Am. 44, 1735–1736.

Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., Evans, A.C., 1996. A unified statistical approach for determining significant voxels in images of cerebral activation. Hum. Brain Mapp. 4, 58–73.

Yousry, T.A., Schmid, U.D., Alkadhi, H., Schmidt, D., Peraud, A., Buettner, A., Winkler, P., 1997. Localization of the motor hand area to a knob on the precentral gyrus. A new landmark. Brain 120, 141–157.