NeuroImage 101 (2014) 364–369

Contents lists available at ScienceDirect

NeuroImage

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

Superficial amygdala and hippocampal activity during affective music listening observed at 3 T but not 1.5 T fMRI

Stavros Skouras [a,c,*], Marcus Gray [b,d], Hugo Critchley [b], Stefan Koelsch [a,c]

[a] Cluster of Excellence, Languages of Emotion, Freie Universität Berlin, Germany
[b] Clinical Imaging Sciences Centre, Falmer, UK
[c] Department of Psychology, University of Sussex, Falmer, UK
[d] Centre for Advanced Imaging, University of Queensland, St Lucia, Australia

ARTICLE INFO

Article history:
Accepted 5 July 2014
Available online 12 July 2014

Keywords:
3 T
1.5 T
Amygdala
Hippocampus
Cuneus
Auditory
Emotion
Joy
Fear
Music

ABSTRACT

The purpose of this study was to compare 3 T and 1.5 T fMRI results during emotional music listening. Stimuli comprised of psychoacoustically balanced instrumental musical pieces, with three different affective expressions (fear, neutral, joy). Participants (N = 32) were split into two groups, one subjected to fMRI scanning using 3 T and another group scanned using 1.5 T. Whole brain t-tests (corrected for multiple comparisons) compared joy and fear in each of the two groups. The 3 T group showed significant activity differences between joy and fear localized in bilateral superficial amygdala, bilateral hippocampus and bilateral auditory cortex. The 1.5 T group showed significant activity differences between joy and fear localized in bilateral auditory cortex and cuneus. This is the first study to compare results obtained under different field strengths with regard to affective processes elicited by means of auditory/musical stimulation. The findings raise concern over false negatives in the superficial amygdala and hippocampus in affective studies conducted under 1.5 T and caution that imaging improvements due to increasing magnetic field strength can be influenced by region-specific characteristics.

Introduction

Over the past decade, fMRI scanning at 3 T has evolved from being a promising new technology to being a standard technology used in cognitive neuroscientific studies. Although it is generally accepted that increasing magnetic field strength has led to improved neuroimaging quality there is certain controversy surrounding the matter, as the comparison with the 1.5 T predecessor is not trivial nor well documented (Wardlaw et al., 2012). Moreover, there exist no published studies comparing field strength effects with regard to affective processes elicited by means of auditory stimulation, because the vast majority of published studies are focused on differences during sensory and motor tasks.

As an early review suggests, the increase from 1.5 T to 3 T is accompanied by improved detection of activity in terms of both volume and reliability (Voss et al., 2006). This is due to increased signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) which in theory can be exchanged for increased spatial and/or temporal resolution. However, the authors caution that factors other than field strength might underlie these effects (Voss et al., 2006). A more recent review takes an opposing stance, implying that the merits of 3 T scanning are exaggerated, particularly because the actual SNR improvement from 1.5 T to 3 T is only 25% of what was theoretically expected (Wardlaw et al., 2012). Furthermore, there seem to be more artefacts occurring at 3 T, particularly in regions near the base of the skull. The authors also caution that other factors, which are uncontrolled across studies, may be influencing the overall pattern of observations and that their review is compromised by underspecified report practises characterizing the literature on the subject (Wardlaw et al., 2012).

Relatively few studies have made explicit comparisons between 1.5 T and 3 T using human subjects. One of these studies used no experimental task (Triantafyllou et al., 2005), three studies used motor tasks (Fera et al., 2003; Hoenig et al., 2005; Yongbi et al., 2001), one study used a visual task (Okada et al., 2005) and one study used sensorimotor (visual) stimulation (Krüger et al., 2001). Only one study used a cognitive task (Hoenig et al., 2005), only two studies used auditory tasks (Han and Talavage, 2011; Rabe et al., 2006) and only one used an affective (visual) task (Krasnow et al., 2003).

The only published study making comparisons between 1.5 T and 3 T while utilizing cognitive rather than sensory or motor tasks (Hoenig et al., 2005), suggests that the improved sensitivity associated with imaging at 3 T can lead beyond the identification of stronger activity in a particular region, to the identification of additional activity in regions that are not observable when imaging at 1.5 T. Results from the only

* Corresponding author at: Freie Universität Berlin, Cluster of Excellence: Languages of Emotion, Habelschwerdter Allee 45, 14195 Berlin, Germany. Fax: +49 30 83852887.
E-mail address: mail@stavros-skouras.com (S. Skouras).

http://dx.doi.org/10.1016/j.neuroimage.2014.07.007
1053-8119/

related comparative study to date utilizing an affective task (Krasnow et al., 2003) suggest that imaging at 3 T is also preferable for subcortical regions. That study limited the analysis of the data from the affective task to the amygdala and found a non-significant trend for stronger amygdala activity at 3 T relative to 1.5 T (Krasnow et al., 2003). Paradoxically, higher signal drop-out in the amygdala under 3 T was also observed although this finding did not reach significance (Krasnow et al., 2003). The interested reader should note that susceptibility artefacts and signal drop-out due to 3 T imaging can be mitigated by the use of spiral in-out acquisition sequences (Preston et al., 2004).

Only two prior studies related to field strength have used an auditory task. The first one investigated whether habituation effects during auditory sine-wave stimulation differ between 1.5 T and 3 T and found only marginal, non-significant differences (Rabe et al., 2006). The second study utilized a task involving passively listening to 750 ms of an auditory stimulus every 12 or 24 s and made no explicit comparison between the two field strengths, focusing instead on showing that increasing the number of fMRI datasets by combining 1.5 T and 3 T datasets is preferable to using a fewer number of only 3 T datasets (Han and Talavage, 2011).

Although one previous study had compared data obtained at different magnetic field strengths in relation to emotional effects elicited through visual stimulation (Krasnow et al., 2003), no previous study compared fMRI data obtained at different magnetic field strengths in relation to emotional effects elicited through auditory stimulation. Musical stimulation is particularly suited for eliciting emotions and activity within reward networks (Koelsch, 2014). Even though fMRI noise can confound the undisturbed acoustic perception and emotional appreciation of music (Skouras et al., 2013), its effects are mitigated by the use of earplugs and special soundproof headphones isolating external sounds.

In the present study we use emotional music to address whether there is any noticeable difference in the detectability of emotion-related contrasts, as established with 3 T fMRI. Our recent studies using similar stimulus material show that during music-elicited joy, the superficial amygdala (SF) is the structure with the strongest increase in activity (Koelsch et al., 2013) as well as centrality (Koelsch and Skouras, 2013). We aim to test whether the detectability of this structure is enhanced or diminished when scanning at 1.5 T.

Material and Methods

Participants

32 individuals (aged 19–28 years, M = 22.93, SD = 2.75, 18 females) took part in the experiment. All participants had normal hearing (as assessed with standard pure tone audiometry) and were right-handed (according to self-report). None of the participants was a professional musician or a music student; 16 participants had no or only minimal formal musical training and 16 participants were amateur musicians who had learned at least one musical instrument (mean duration of formal training was 2.78 years). The participants were split into two groups of 16 subjects each; a group that underwent 3 T scanning (aged 19–28 years, M = 22.94, SD = 2.72, 9 females, mean duration of formal training 3.62 years, SD = 5.04) and a group that underwent 1.5 T scanning (aged 19–28 years, M = 22.92, SD = 2.90, 9 females, mean duration of formal training 1.94 years, SD = 2.35).

Independent samples t-tests showed that the two groups did not differ with regard to age ($p = 0.98$) and formal musical training ($p = 0.24$). Exclusion criteria were left-handedness, professional musicianship, a score on Beck's Depression Inventory (Beck and Steer, 1993) of 13 or more points, consumption of alcohol or caffeine exceeding one litre during the 24 h prior to testing, poor sleep during the previous night, past diagnosis of a neurological or psychiatric disorder, and abnormal brain anatomy. All subjects gave written informed consents. The study was conducted according to the Declaration of Helsinki and approved by the ethics committee of the School of Life Sciences and the Psychology Department of the University of Sussex.

Stimuli

Musical stimuli were selected to evoke (a) feelings of joy, (b) feelings of fear, or (c) neither joy nor fear (henceforth referred to as neutral stimuli). There were n = 8 stimuli per category. The stimuli utilized have been used and described in detail previously (Koelsch et al., 2013). Musical stimuli evoked the intended feelings in the sample studied (see Behavioural data analysis). Joyful stimuli consisted of CD-recorded pieces from various epochs and styles (classical music, Irish jigs, jazz, reggae, South American and Balkan music). Fearful musical stimuli were excerpts from soundtracks of suspense movies and video games. To increase the fear-evoking effect of the fear stimuli, their relatively high acoustic roughness was further increased: from each fear excerpt, two copies were obtained and pitch-shifted, one copy was shifted one semitone higher, the other copy a tritone lower (see also Fritz et al., 2009; Koelsch et al., 2006). Then, all three versions of one excerpt (original pitch, one semitone higher, and a tritone lower) were rendered as a single wav-file (pitch-shift and rendering was performed using Ableton Live, version 8.0.4, Ableton AG, Berlin, Germany). The complete list of joyful and fearful stimuli is provided in Supplementary Table 1. Neutral stimuli were created using the MIDI toolbox for Matlab (Eerola and Toiviainen, 2004) and Ableton Live (version 8.0.4, Ableton Inc., New York, USA). Neutral stimuli comprised of sequences of isochronous tones, for which the pitch classes were randomly selected from a pentatonic scale and the timbres where chosen from a high quality natural instrument library (X-Sample Chamber Ensemble, Winkler, Stahl GbR, Detmold, Germany). Stimuli were edited using Praat (version 5.0.29; Boersma, 2002), so that they all had the same length (30 s), 1.5 s fade-in/fade-out ramps, and the same RMS power. All stimuli were matched across conditions in triplets (joy-neutral-fear) with regard to tempo (beats per minute), mean F0 pitch, F0 pitch variation, pitch centroid value, spectral complexity, and spectral flux. This was confirmed by an acoustic analysis of the stimuli using 'Essentia', an in-house library for extracting audio and music features from audio files (http://mtg.upf.edu/technologies/essentia). Stimuli tempi ranged between 93 and 149 beats per minute (mean = 117.25, S.D. = 21.70) and fundamental frequencies ranged between 110 and 2637 Hz. The Essentia software was also used to test for differences between stimuli with regard to another 177 acoustical factors. Ten psychoacoustic factors were found to differ significantly between experimental conditions ($p < 0.001$, corrected for multiple comparisons). These factors were mean and variance of F0 salience (reflecting differences in the amount of harmonies and percussive sounds present in the different conditions), mean and variance of sensory dissonance, mean chord strength and key strength (reflecting the discriminability of these tonal features) the mean and variance of spectral flux (a measure of spectral variation within sounds), mean spectral crest (a measure of the inhomogeneity, or noisiness, of the spectrum) and mean spectral complexity (which correlates with the amount of different timbres that are present in a piece). For additional details see Koelsch et al., 2013. The values of these factors associated with each stimulus were used in the fMRI data analysis as additional regressors of the general linear model's (GLM) design matrix (see also Data Analysis).

Procedure

Prior to the MRI session, participants were presented with short (12 s) versions of each stimulus to obtain familiarity ratings (for the importance of this see Pereira et al., 2011). Participants rated their familiarity with each piece on a four-point scale (ranging from “To my knowledge I have never heard this piece before”, to “I know this piece, and I know who composed, or performed it”). Familiarity ratings from six participants who stated at least once that they knew the name of