Neuroscience Letters 593 (2015) 35–39

Contents lists available at ScienceDirect

Neuroscience Letters

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

Research article

The specificity of neural responses to music and their relation to voice processing: An fMRI-adaptation study

Jorge L. Armony[a,b,c,d,*], William Aubé[a,b,c], Arafat Angulo-Perkins[e], Isabelle Peretz[a,b,c], Luis Concha[a,e]

[a] International Laboratory for Brain, Music and Sound Research (BRAMS), Montreal, Canada
[b] Centre for Research on Brain, Language and Music (CRBLM), Montreal, Canada
[c] Department of Psychology, Université de Montréal, Montreal, Canada
[d] Douglas Mental Health University Institute and Department of Psychiatry, McGill University, Montreal, Canada
[e] Universidad Nacional Autónoma de México, Queretaro, Mexico

HIGHLIGHTS

- We investigated neural specificity of brain responses to musical stimuli.
- Modality-specific adaptation occurs in visual and auditory cortical regions.
- Music elicits stronger responses than voice in the anterior superior temporal gyrus.
- A region in anterior superior temporal gyrus displays music-specific adaptation.
- Our findings support the existence of music-preferred neurons.

ARTICLE INFO

Article history:
Received 14 November 2014
Received in revised form 3 February 2015
Accepted 6 March 2015
Available online 10 March 2015

Keywords:
Functional magnetic resonance imaging
Music perception
Voice processing
Repetition suppression
fMRI adaptation
Musicianship

ABSTRACT

Several studies have identified, using functional magnetic resonance imaging (fMRI), a region within the superior temporal gyrus that preferentially responds to musical stimuli. However, in most cases, significant responses to other complex stimuli, particularly human voice, were also observed. Thus, it remains unknown if the same neurons respond to both stimulus types, albeit with different strengths, or whether the responses observed with fMRI are generated by distinct, overlapping neural populations. To address this question, we conducted an fMRI experiment in which short music excerpts and human vocalizations were presented in a pseudo-random order. Critically, we performed an adaptation-based analysis in which responses to the stimuli were analyzed taking into account the category of the preceding stimulus. Our results confirm the presence of a region in the anterior STG that responds more strongly to music than voice. Moreover, we found a music-specific adaptation effect in this area, consistent with the existence of music-preferred neurons. Lack of differences between musicians and non-musicians argues against an expertise effect. These findings provide further support for neural separability between music and speech within the temporal lobe.

1. Introduction

A growing neuroimaging literature is providing strong support for the notion that, similar to what has been reported in the visual domain, a spatial segregation of responses to different complex auditory stimuli exists along the temporal lobes. For instance, several studies have reported that the superior temporal sulcus (STS) appears to preferentially respond to human voices, compared to nonvocal sounds or vocalizations from other animals [1,2]. More recently, several fMRI studies [3–6] have identified a region within the anterior superior temporal gyrus (STG) that shows enhanced responses to musical stimuli compared to human voice, including speech, and other complex acoustic stimuli. These studies have rekindled the old debate on the relation between language and music, particularly in terms of their evolutionary origins [7]; in particular, they further support to the proposal that, consistent with some lesion studies showing differential language- or music-specific deficits [7,8], there are neural networks specific for the processing of music. Nonetheless, these previous studies need to be interpreted with caution, as in most cases, the regions

* Corresponding author at: Douglas Institute – Research 6875 LaSalle boulevard Verdun, QC H4H 1R3, Canada. Tel.: +1 514 343 6111x28300.
E-mail address: jorge.armony@mcgill.ca (J.L. Armony).

http://dx.doi.org/10.1016/j.neulet.2015.03.011
0304-3940/

Table 1
Differences (in absolute value) in the main acoustic parameters between successive auditory stimuli for the different categories of interest.

Absolute differences in acoustic parameters: mean (SD)								
Category	Duration (ms)	Spectral centroid (Hz)	Spectral flux (a.u.)	Intensity flux (a.u.)	Spectral regularity (a.u.)	RMS (a.u.)	HNR (dB)	Median f0 (Hz)
M_M	184 (157)	367.3 (367.7)	9.6 (8.4)	52.3 (44.0)	0.41 (0.31)	0.02 (0.01)	7.89 (5.03)	55.0 (48.1)
M_V	334 (203)	1003.1* (643.0)	34.4 (30.9)	118.9 (82.6)	0.44 (0.28)	0.08* (0.07)	7.09 (5.31)	133.5 (81.1)
V_V	499* (260)	552.1 (459.6)	33.1* (27.4)	84.0 (86.1)	0.51 (0.45)	0.09* (0.08)	4.46 (4.24)	96.5 (55.8)
V_M	332 (177)	742.5 (364.3)	44.8* (38.1)	95.0 (56.0)	0.53 (0.31)	0.11* (0.09)	5.37 (3.51)	94.5 (75.9)

The spectral centroid (weighted mean of spectrum energy) reflects the global spectral distribution and has been used to describe the timber, whereas the spectral flux conveys spectrotemporal information (variation of the spectrum over time) and the spectral regularity represents the degree of uniformity of the successive peaks of the spectrum [25]. The intensity flux is a measure of loudness as a function of time [26]. Other measures were also computed such as the root mean square (RMS), the harmonic to noise ratio (HNR) and the median f0 [11].
M_M: music preceded by music; M_V: music preceded by voice; V_V: voice preceded by voice; V_M: voice preceded by music; a.u.: arbitrary units.
* Significantly different from M_M ($p < 0.05$, Bonferroni corrected).

found to respond preferentially to music stimuli also respond to human vocalizations, albeit to a lesser degree than to music [9]. The question remains as to whether the same neurons respond to both types of stimulus, to a different degree or, alternatively, the observed responses are generated by spatially overlapping, yet distinct groups of neurons. Unfortunately, the nature of the BOLD signal measured with fMRI, and its limited spatial resolution, do not allow for directly identifying which neurons are active in response to a given stimulus. However, we can take advantage of the nonlinear dynamics of neural activity to indirectly address this question, by using the so-called fMRI adaptation paradigm [10]. This approach, based on the principle of neuronal adaptation/habituation, reflects the fact that the observed BOLD signal to successive stimuli depends on whether these engage the same group of neurons or not. That is, the activity associated with two stimuli will get smaller with repetition if they activate the same neuronal pool than if they stimulate different neurons. Critically, although adaptation is strongest when repeating the same stimulus, it can also be observed when different exemplars from the same category are presented, and can thus be used to identify those brain regions in which different types of stimuli share a common neural representation.

Here, we employed this strategy to explore the specificity of the neural responses in the previously identified "music area". Specifically, we performed a new analysis of a previous experiment [11] by grouping individual stimuli not only based on the category to which they belonged, but also taking into account the one that preceded them.

2. Methods

The study presented here constitutes a new analysis of an experiment designed to identify brain responses to faces, nonlinguistic vocalizations and musical excerpts expressing different emotions. Results from the main analysis (i.e., as a function of stimulus category and emotion), as well as a more detailed description of the experimental paradigm and stimulus characteristics are presented elsewhere [11].

2.1. Paradigm

Forty-seven healthy right-handed volunteers (20 female, mean age: 26.4) with no history of hearing impairments participated in the study.

Stimuli consisted of novel musical clips (played with piano or violin) [11–13], nonlinguistic vocalizations [14,15] and faces [16,17]. Sixty different exemplars of each category were presented in a pseudo-random order (with no more than 3 stimuli of the same category presented consecutively). Table 1 shows the mean difference (absolute value) in the magnitude of several relevant acoustic parameters [11] between each auditory stimulus and its preceding one, as a function of their category. Stimulus duration was on average 1.5 s with a mean intertrial interval (ITI) of 2.5 s. Participants' task was to detect the sporadic presentation of a visual (inverted face) or auditory (500 Hz pure tone) target.

2.3. Image acquisition

Functional images were acquired in a 3 T MR750 scanner (General Electric, Wuaukesha, Wisconsin) with a 32-channel coil using parallel imaging with an acceleration factor of 2 (FOV = $256 \times 256\,mm^2$, matrix = 128×128, TR = 3 s, TE = 40 ms; voxel size = $2 \times 2 \times 3\,mm^3$). A 3D T_1-weighted image was also acquired and used for registration (voxel size = $1 \times 1 \times 1\,mm^3$, TR = 2.3 s, TE = 3 ms).

2.4. Statistical analysis

Image preprocessing and analysis were carried out using SPM8 (Wellcome Trust Centre for Neuroimaging, London, United Kingdom; http://www.fil.ion.ucl.ac.uk/spm). Trials were assigned to different conditions depending on the category of the stimulus (face, music or vocalization) and that of the immediately preceding stimulus, resulting in nine conditions of interest (F_F: face preceded by face; F_V: face preceded by voice; F_M: face preceded by music; V_F: voice preceded by face; V_V: voice preceded by voice; V_M: voice preceded by music; M_F: music preceded by face; M_V: music preceded by voice and M_M: music preceded by music). Additional categories of no interest consisted of those corresponding to visual or auditory target and to experimental stimuli that followed a target or a null event (there were too few instances of these conditions for a meaningful analysis). Each event was modeled as a boxcar of a length equal to the duration of the stimulus presentation, convolved with the canonical hemodynamic response function. The six movement parameters obtained in the realignment procedure were also included in the model as nuisance regressors. Parameter estimates for the nine conditions of interest obtained in the first-level, single-subject analysis were taken to a second-level repeated-measures ANOVA. Adaptation effects were tested by contrasting stimuli belonging to the same category, say faces, but which were preceded by either a stimulus of the same category (i.e., faces) or of a different one (i.e., vocalizations or music).

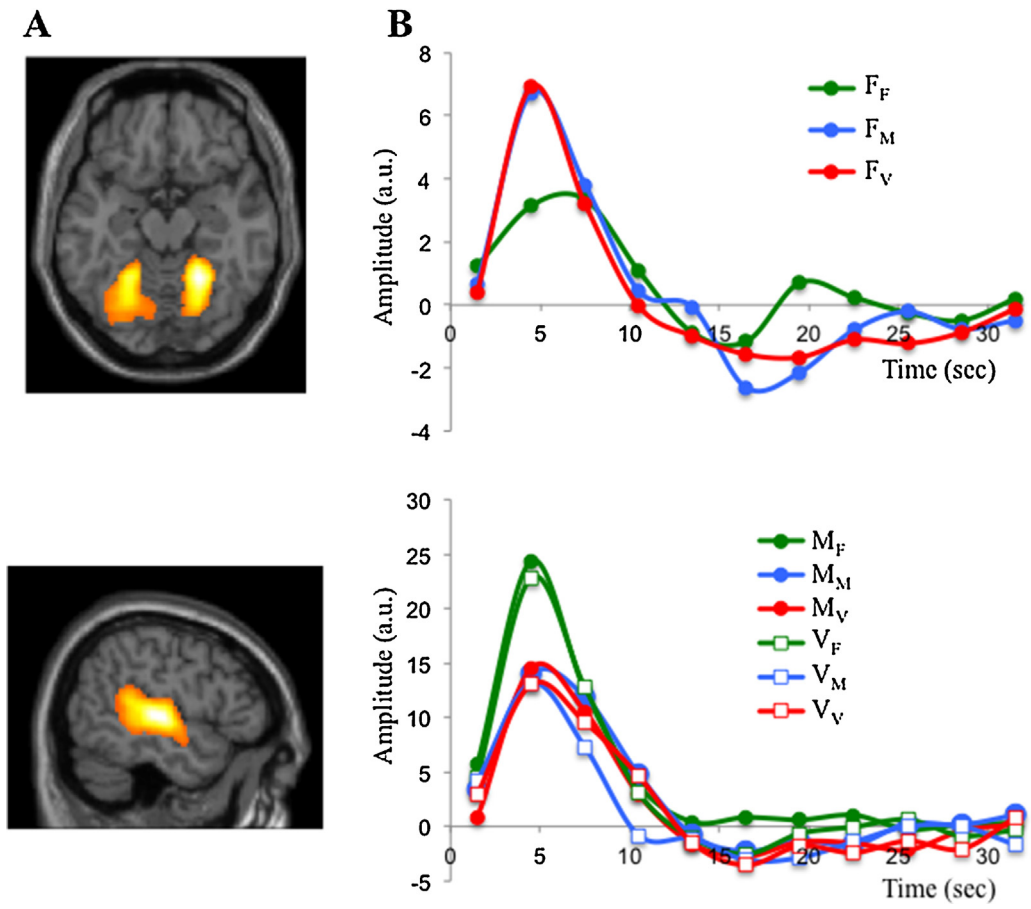

Fig. 1. Modality-specific adaptation. (A) Clusters in the fusiform gyrus (*Top*) and Temporal Lobe (*Bottom*) in which significant within-modality adaptation effects were observed. (B) Group-averaged peri-stimulus time histograms (PSTHs) for the different conditions for the two clusters with peaks in [22 –52 –16] and [–50 –24 4], respectively. F_F: face preceded by face; F_V: face preceded by voice; F_M: face preceded by music; M_F: music preceded by face; M_M: music preceded by music; M_V: music preceded by voice; V_F: voice preceded by face; V_M: voice preceded by music; V_V: voice preceded by voice.

We further confirmed the reliability of the results obtained by conducting a leave-one-subject-out (LOSO) cross-validation analysis [18]. Namely, we performed 47 separate group analyses with 46 subjects, each time leaving a different subject out. For each of these analyses, a cluster of voxels showing significant adaptation to music ($p < 0.05$, uncorrected) was identified, serving as an independent "region-of-interest" (ROI) for the subject left out. Mean parameter estimates for these LOSO ROIs were extracted for the left-out subject and post-hoc group analysis conducted.

3. Results

3.1. Modality-specific adaptation

A different- vs. same-modality comparison confirmed the adaptation effects in sensory regions. Specifically, the contrast of responses to visual stimuli (faces) that were preceded by auditory stimuli (vocalizations or music) minus those preceded by faces, $[(F_M + F_V)/2 - F_F]$, yielded significant activations in visual areas, including the fusiform gyrus ($[x\ y\ z] = [22\ -52\ -16]$, $Z = 7.61$ and $[x\ y\ z] = [-22\ -52\ -12]$, $Z = 6.19$; see Fig. 1A). Similarly, adaptation effects were observed in auditory regions of the temporal lobes for auditory stimuli that were preceded by auditory stimuli (either of the same or different categories, i.e., music or vocalizations), compared to the case where they were preceded by visual stimuli ($[x\ y\ z] = [52\ -20\ 2]$, $Z = 10.97$ and $[x\ y\ z] = [-50\ -24\ 4]$, $Z = 11.41$; see Fig. 1B).

3.2. Music adaptation

The contrast music-preceded-by-voice minus music-preceded-by-music ($[M_V - M_M]$; i.e., corresponding to the music-specific adaptation effects), masked inclusively by the main contrast music minus voice, revealed a cluster in the right superior temporal gyrus ($[x\ y\ z] = [46\ -12\ 2]$; Fig. 2A). As shown in Fig. 2B, the extracted parameter estimates, as well as the BOLD signal (Fig. 2C), show a significant adaptation effect for music: responses to musical stimuli were smaller when these were immediately preceded by other musical stimuli, as compared to when the previous stimulus was either faces [$t(46) = 8.7$, $p < 0.001$] or vocalizations [$t(46) = 2.3$, $p = 0.01$]. This adaptation effect was also confirmed by a leave-one-subject-out cross-validation procedure [$p = 0.02$, mean cluster size = 70; see Methods) and the prevalence map, shown in Fig. 2D.

4. Discussion

As expected, we observed sensory-specific adaptation effects in cortical areas; that is, responses to faces in visual areas, including occipital cortex and fusiform gyrus, were reduced when a face was preceded by another face, compared to when an auditory stimuli was presented before it. Likewise, activity in auditory regions in the temporal lobe was modulated by the sensory modality of the previous stimulus. Importantly, the adaptation to auditory stimuli in these areas was similar whether stimulus pair belonged to the same category (i.e., music/music or voice/voice) or not (i.e., music/voice or voice/music). These results are consistent with the fact that

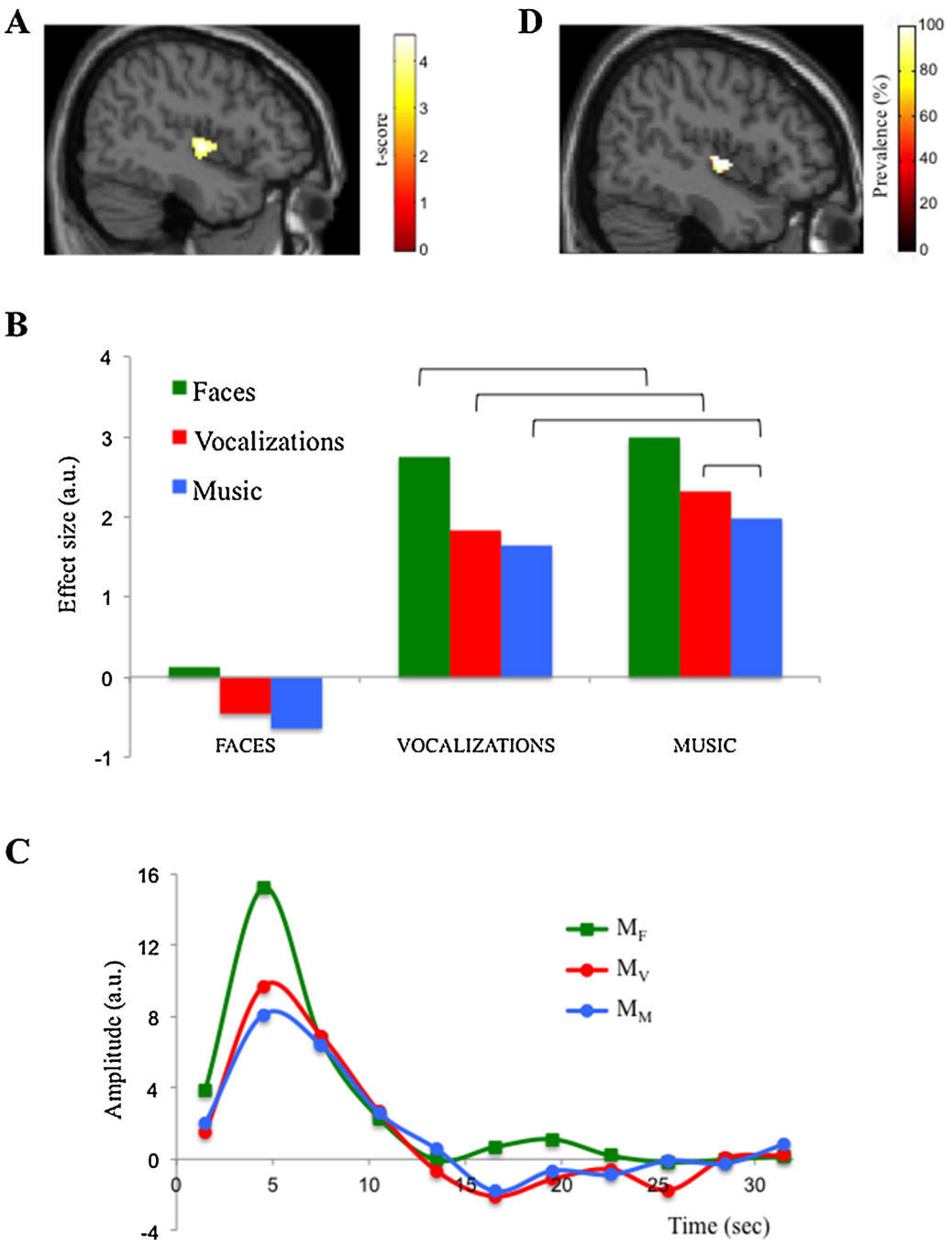

Fig. 2. Adaptation to musical stimuli. (A) Statistical parametric map (SPM) showing the cluster within the Superior Temporal Gyrus in which there was a significant adaptation to musical excerpts. (B) Group- and cluster-averaged effect sizes for the conditions of interest (brackets indicate significant differences between conditions). (C) Group-averaged peri-stimulus time histograms (PSTHs) showing responses to music stimuli as a function of the category of the stimuli that preceded them. (D) Sagittal section showing the percentage of individual Leave-One-Subject-Out (LOSO) analyses (one per subject) that yielded a significant music adaptation effect (see Methods section for details).
M_F: music preceded by face; M_V: music preceded by voice; M_M: music preceded by music.

neurons in primary auditory regions respond to basic acoustic features across a large variety of stimulus classes, including human voice and music [5].

In contrast, the significantly stronger adaptation effect for music stimuli when they were preceded by other (different) music stimuli than by vocalizations observed in a region within the STG provides strong support for the existence of "music-preferring" neurons in this area, which was previously identified to respond more strongly to music than to other complex auditory stimuli, including human voice [3–6]. Nonetheless, adaptation was more pronounced for voice-music than face-music pairs, suggesting some degree of neuronal sharing in the representation of these two types of human-produced stimuli. Finally, the similar adaptation effects for responses elicited by vocalizations when preceded by either other vocalizations or music points against the presence of voice-preferred neurons in this region.

4.1. Limitations

There are a number of limitations that need to be considered when interpreting the results presented here. First, although fMRI adaptation is being increasingly used to explore the neural

specificity of brain responses, its precise neuronal substrates are still unknown [19]. Nonetheless, it is important to point out that, unlike some of the early fMRI adaptation work employing blocked designs, our results cannot be due to stimulus expectation, as the transition probabilities were similar across conditions and therefore could not be predicted [20].

Secondly, we used a restricted set of music and vocal stimuli. While this restriction was partly by design, as it allowed us a better control of the acoustic properties of the stimuli, it does limit the generalizability of our findings to other types of auditory information, particularly speech, as well as more complex musical stimuli. Additionally, differences between successive stimuli in some basic acoustic parameters significantly differed among conditions, as shown in Table 1. Although none of these parameters seemed to explain on its own the pattern of activation shown in Fig. 2B, we cannot rule out the possibility that a combination of them, or another one not measured here, could underlie, at least in part, the adaptation effects obtained in this experiment. Future studies employing a larger variety of stimuli and a more thorough control of their acoustic parameters should help answer this question.

Finally, our study does not allow us to address the important question of whether the music-preferring neurons represent an innate, hardwire system, possibly arising from an "invasion" or "recycling" [21] of the language system [22], as all our participants were exposed to music during their life and thus we cannot rule out some degree of experience-dependent plasticity [23,24]. However, the fact that we found no significant differences between professional musicians and non-musicians makes it unlikely that the effects observed are solely due to learning.

5. Conclusion

The results presented here, employing an fMRI-adaptation analysis, suggest the existence of music-specific, or at least preferred, neurons in the previously described "music area" located in the anterior superior temporal gyrus. Nonetheless, our findings also indicate some degree of neural sharing in the representation of social information conveyed by different means, namely voice and music.

Acknowledgments

This work was partly funded by grants from the National Science and Engineering Research Council of Canada (NSERC, 262439-2009) and the Canadian Institutes of Health Research (CIHR, MOP-93762) to JLA, and from Consejo Nacional de Ciencia y Tecnología de México (CONACyT, IE252-120295) and Universidad Nacional Autónoma de México (UNAM-DGAPA, IN212811) to LC. APA and WA were supported by CONACyT and NSERC graduate fellowships, respectively.

References

[1] P. Belin, R.J. Zatorre, P. Lafaille, P. Ahad, B. Pike, Voice-selective areas in human auditory cortex, Nature 403 (2000) 309–312.
[2] S. Fecteau, J.L. Armony, Y. Joanette, P. Belin, Is voice processing species-specific in human auditory cortex? An fMRI study, Neuroimage 23 (2004) 840–848.
[3] A. Angulo-Perkins, W. Aube, I. Peretz, F.A. Barrios, J.L. Armony, L. Concha, Music listening engages specific cortical regions within the temporal lobes: differences between musicians and non-musicians, Cortex 59C (2014) 126–137.
[4] E. Fedorenko, J.H. McDermott, S. Norman-Haignere, N. Kanwisher, Sensitivity to musical structure in the human brain, J. Neurophysiol. 108 (2012) 3289–3300.
[5] A.M. Leaver, J.P. Rauschecker, Cortical representation of natural complex sounds: effects of acoustic features and auditory object category, J. Neurosci. 30 (2010) 7604–7612.
[6] C. Rogalsky, F. Rong, K. Saberi, G. Hickok, Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging, J. Neurosci. 31 (2011) 3843–3852.
[7] I. Peretz, The nature of music from a biological perspective, Cognition 100 (2006) 1–32.
[8] I. Peretz, Music, language, and modularity framed in action, Psychol. Belg. 49 (2009) 157–175.
[9] I. Peretz, D. Vuvan, M.-E. Lagrois, J.L. Armony, Neural Overlap in Processing Music and Speech, Philos. Trans. R. Soc. London Ser. B Biol. Sci. 370 (2015), 2014009.
[10] K. Grill-Spector, R. Malach, fMR-adaptation: a tool for studying the functional properties of human cortical neurons, Acta Psychol. (Amst.) 107 (2001) 293–321.
[11] W. Aube, A. Angulo-Perkins, I. Peretz, L. Concha, J.L. Armony, Fear across the senses: brain responses to music, vocalizations and facial expressions, Soc. Cogn. Affect Neurosci. 10 (2015) 399–407.
[12] W. Aube, I. Peretz, J.L. Armony, The effects of emotion on memory for music and vocalisations, Memory 21 (2013) 981–990.
[13] S. Vieillard, I. Peretz, N. Gosselin, S. Khalfa, L. Gagnon, B. Bouchard, Happy, sad, scary, and peaceful musical excerpts for research on emotions, Cognit. Emot. 22 (2008) 720–752.
[14] J.L. Armony, C. Chochol, S. Fecteau, P. Belin, Laugh (or cry) and you will be remembered: influence of emotional expression on memory for vocalizations, Psychol. Sci. 18 (2007) 1027–1029.
[15] S. Fecteau, P. Belin, Y. Joanette, J.L. Armony, Amygdala responses to nonlinguistic emotional vocalizations, Neuroimage 36 (2007) 480–487.
[16] K. Sergerie, M. Lepage, J.L. Armony, A process-specific functional dissociation of the amygdala in emotional memory, J. Cogn. Neurosci. 18 (2006) 1359–1367.
[17] K. Sergerie, M. Lepage, J.L. Armony, Influence of emotional expression on memory recognition bias: a functional magnetic resonance imaging study, Biol. Psychiatry 62 (2007) 1126–1133.
[18] M. Esterman, B.J. Tamber-Rosenau, Y.C. Chiu, S. Yantis, Avoiding non-independence in fMRI data analysis: leave one subject out, Neuroimage 50 (2010) 572–576.
[19] U. Noppeney, Characterization of multisensory integration with fMRI: experimental design, statistical analysis, and interpretation, in: M.M. Murray, M.T. Wallace (Eds.), The Neural Bases of Multisensory Processes, CRC press, Boca Raton (FL), 2012.
[20] C. Summerfield, E.H. Trittschuh, J.M. Monti, M.M. Mesulam, T. Egner, Neural repetition suppression reflects fulfilled perceptual expectations, Nat. Neurosci. 11 (2008) 1004–1006.
[21] S. Dehaene, L. Cohen, Cultural recycling of cortical maps, Neuron 56 (2007) 384–398.
[22] I. Peretz, W. Aube, J.L. Armony, Towards a neurobiology of musical emotions, in: E. Altenmüller, S. Schmidt, E. Zimmermann (Eds.), The Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man, Oxford University Press, Oxford, UK, 2013.
[23] L.J. Trainor, A. Shahin, L.E. Roberts, Effects of musical training on the auditory cortex in children, Ann. N. Y. Acad. Sci. 999 (2003) 506–513.
[24] S.C. Herholz, R.J. Zatorre, Musical training as a framework for brain plasticity: behavior, function, and structure, Neuron 76 (2012) 486–502.
[25] J. Marozeau, A. de Cheveigneí, S. McAdams, S. Winsberg, The dependency of timbre on fundamental frequency, J. Acoust. Soc. Am. 114 (2003) 2946–2957.
[26] B.R. Glasberg, B.C.J. Moore, A model of loudness applicable to time-varying sounds, J. Audio Eng. Soc. 50 (2002) 331–342.