UNIL | Université de Lausanne

Unicentre
CH-1015 Lausanne
http://serval.unil.ch

*Year :* 2019

# IMAGING EMOTIONAL SOUNDS PROCESSING AT 7T

Grisendi Tiffany

Grisendi Tiffany, 2019, IMAGING EMOTIONAL SOUNDS PROCESSING AT 7T

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_C1008758C9014

**Droits d'auteur**
L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

**Copyright**
The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

**Département des Neurosciences Cliniques**

# IMAGING EMOTIONAL SOUNDS PROCESSING AT 7T

## Thèse de doctorat en Neurosciences

présentée à la

Faculté de Biologie et de Médecine
de l'Université de Lausanne

par

## Tiffany Grisendi

Ingénieure en Sciences et Technologies du Vivant de l'Ecole Polytechnique Fédérale de Lausanne, Suisse

## Jury

Prof. Jean-Pierre Hornung, Président
Prof. Stephanie Clarke, Directrice
Dr. Sandra Da Costa, Co-Directrice
Prof. Sascha Frühholz, Expert
Prof. Patric Hagmann, Expert

Thèse n° 250

**Lausanne 2019**

***Programme doctoral interuniversitaire en Neurosciences des Universités de Lausanne et  Genève***

LEMANIC NEURO SCIENCE DOCTORAL SCHOOL
UNIL | Université de Lausanne
UNIVERSITÉ DE GENÈVE

**UNIL** | Université de Lausanne

**UNIVERSITÉ DE GENÈVE**

*Programme doctoral interuniversitaire en Neurosciences des Universités de Lausanne et Genève*

# Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

| | | | |
|---|---|---|---|
| **Président·e** | Monsieur | Prof. | Jean-Pierre Hornung |
| **Directeur·trice de thèse** | Madame | Prof. | Stephanie Clarke |
| **Co-directeur·trice de thèse** | Madame | Dre. | Sandra Da Costa |
| **Expert·e·s** | Monsieur | Prof. | Sascha Frühholz |
| | Monsieur | Prof. | Patric Hagmann |

le Conseil de Faculté autorise l'impression de la thèse de

**Madame Tiffany Grisendi**

Ingénieure en Sciences et Technologies du Vivant
de l'Ecole Polytechnique Fédérale de Lausanne

intitulée

**Imaging Emotional Sounds Processing AT 7T**

Lausanne, le 15 juillet 2019

pour Le Doyen
de la Faculté de Biologie et de Médecine

Prof. Jean-Pierre Hornung

# Acknowledgements

First, I would like to thank all who kindly participated in these studies, and made this work possible.

I would like to express my sincerest gratitude to my thesis director, Professor Stephanie Clarke, for giving me the opportunity to work in an interdisciplinary research group. Her continuous support and the sharing of her outstanding scientific knowledge made me a more experienced researcher. Many thanks to Dr. Sandra Da Costa who became a precious supervisor. I am very grateful to her for her priceless help, for knowledge she was so generous to transmit and especially for the great moments we shared over a coffee or in front of the scanner.

Aside from my thesis directors, I would like to thank Professor Sascha Frühholz, Professor Patric Hagmann and Professor Jean-Pierre Hornung for accepting to be part of my jury.

Thanks to all my colleagues for the support, the scientific and non-scientific exchanges and the friendships. It has been a pleasure to work with all of them.

I would like to thank all my relatives, my family, my step-family and my friends. They always supported me until the end of this thesis. Thanks to my parents who gave me the opportunity to pursue the studies I wanted, encouraged me and helped me when needed. A special thought goes out to my grandfather who will not have seen me finishing this thesis, but who always encouraged me in everything I undertook. Last but not least, an enormous thanks to my husband, Adrien, who supported and encouraged me during this whole endeavor. He was always able to motivate me during critical moments and be by my side to share the good times.

Finally, I want to dedicate this thesis to my son, Tilian. The path you must take to reach you goal is not necessarily the one you thought, but you will always reach the summit. Always believe in your dreams whatever they may be.

# Abstract

Emotional sounds and their localization are influential stimuli that we need to process all along our life. Affective information contained in sounds is primordial for the human social communications and interactions. Their accurate localization is important for the identification and reaction to environmental events. This thesis investigate the encoding of emotional sounds within auditory areas and the amygdala (AMY) using 7 Tesla fMRI.

In a first experiment, we studied the encoding of emotion and vocalization and their integration in early-stage auditory areas, the voice area (VA) and the AMY. We described that the response of the early-stage auditory areas was modulated by the vocalization and by the affective content of the sounds, and that this affective modulation is independent of the category of sounds. In contrast, AMY process only the emotional part, while VA is responsible for the processing of the emotional valence specifically for the human vocalization (HV) categories. Finally, we described a functional correlation between VA and AMY in the right hemisphere for the positive vocalizations only.

In a second experiment, we investigated how the spatial origin of an emotional sound (HV or non-vocalizations) modulated its processing within early-stage auditory areas and VA. We highlighted a left hemispace preference for the positive vocalizations encoded bilaterally in the primary auditory cortex (PAC). Moreover, comparison with the first study indicated that the saliency of emotional valence could be increased by spatial cues, but that the encoding of vocalization is not impacted by the spatial context.

Finally, we examined the functional correlations between early-stage auditory areas and VA and how they are modulated by the sound category, the valence and the lateralization. We documented a strong coupling between VA and early-stage auditory areas during the presentation of emotional HV, but not for other environmental sounds. The category of sound modulated strongly the functional correlations between VA, PAC and auditory belt areas, while the spatial positioning induced only a weak modulation and no modulation was caused by the affective content.

Overall, these studies demonstrate that the affective load modulates the processing of sounds within VA only for HV, and that this preference for vocalizations impacts the functional correlations of VA with other auditory regions. This strengthens the importance of VA as a computation hub for the processing of emotional vocalizations.

# Résumé

Les sons émotionnels ainsi que leur localisation sont des stimuli importants que nous devons traiter tout au long de notre vie. L'information affective contenue dans les sons est primordiale pour les communications et interactions sociales. Leur localisation correcte est importante pour l'identification et la réaction par rapport aux événements nous entourant. Cette thèse étudie l'encodage des sons émotionnels dans les aires auditives et l'amygdale (AMY) en utilisant l'IRM fonctionnel à 7 Tesla.

Dans une première expérience, nous avons étudié l'encodage des émotions et des vocalisations, ainsi que leur intégration dans les aires auditives primaires et non-primaires, dans l'aire des voix (VA) et dans AMY. Nous avons décrit que la réponse des aires auditives primaires et non-primaires étaient modulées par les vocalisations ainsi que par le contenu affectif des sons, et que cette modulation affective était indépendante de la catégorie sonore. En revanche, AMY traite uniquement la partie émotionnelle, tandis que la VA est responsable du traitement de la valence émotionnelle spécifiquement pour les vocalisations humaines (HV). Finalement, nous avons décrit une corrélation fonctionnelle entre VA et AMY dans l'hémisphère droit pour les vocalisations positives uniquement.

Dans une seconde expérience, nous avons cherché à comprendre de quelle manière l'origine spatiale d'un son émotionnel (HV et non-vocalisations) modulait son traitement dans les aires auditives, primaires et non-primaires, et VA. Nous avons mis en évidence une préférence de l'hémi-champ gauche pour les vocalisations positive encodées bilatéralement dans le cortex auditif primaire (PAC). De plus, une comparaison avec la première étude a indiqué que l'importance de la valence émotionnelle pourrait être augmentée grâce aux indices spatiaux, mais que l'encodage des vocalisations n'étaient pas impacté par le contexte spatial.

Finalement, nous avons examiné les corrélations fonctionnelles entre les aires auditives primaires, non-primaires et VA afin d'évaluer de quelle manière elles étaient modulées par la catégorie sonore, la valence et la latéralisation. Nous avons mis en évidence un fort couplage entre VA et les aires auditives primaires et non-primaires durant la présentation des HV émotionnelles, mais cet effet n'était pas présent pour les autres sons environnementaux. La catégorie sonore modulait fortement les corrélations fonctionnelles entre VA, PAC et les régions auditives latérales, alors que le positionnement spatial n'influençait que faiblement leur modulation. De plus, il n'y avait pas de modulation causée par le contenu affectif.

En résumé, ces études démontrent que le contenu affectif module le traitement des sons dans VA uniquement pour les HV, et que cette préférence pour les vocalisations a un impact sur les corrélations

fonctionnelles de cette région avec les autres régions auditives. Cela souligne l’importance de VA comme centre computationnel pour le traitement des vocalisations émotionnelles.

# Table of contents

# List of figures

# List of abbreviations

| | |
|---|---|
| A1 | primary auditory area |
| AC | auditory cortex |
| AMY | amygdala |
| ANOVA | analysis of variance |
| ASD | autism spectrum disorder |
| BOLD | blood oxygenation level dependent |
| DCM | dynamic causal modelling |
| dHb | deoxygenated haemoglobin |
| DSI | diffusion spectrum imaging |
| EEG | electroencephalogram |
| EPI | echo planar imaging |
| FDR | false discovery rate |
| fMRI | functional magnetic resonance imaging |
| GLM | general linear model |
| Hb | oxygenated haemoglobin |
| HG | Heschl’s gyrus |
| HRF | Hemodynamic Response Function |
| HV0 | human vocalizations with neutral emotional valence |
| HVN | human vocalizations with negative emotional valence |
| HVP | human vocalizations with positive emotional valence |
| ILD | interaural level difference |
| ITD | interaural time difference |
| LH | left hemisphere |
| MP2RAGE | magnetic prepared 2 rapid acquisition gradient echoes |
| NV0 | non-vocalizations with neutral emotional valence |
| NVN | non-vocalizations with negative emotional valence |

| | |
|---|---|
| NVP | non-vocalizations with positive emotional valence |
| PAC | primary auditory cortex |
| PET | positron emission tomography |
| R | rostral (primary) auditory area |
| RH | right hemisphere |
| ROI | region of interest |
| SNR | signal-to-noise ratio |
| SOC | superior olivary complex |
| STG | superior temporal gyrus |
| STS | superior temporal sulcus |
| TE | echo time |
| TI | inversion time |
| TR | repetition time |
| VA | voice area |

# 1 Introduction

## 1.1 General introduction

Sounds, and in particular emotional sounds are highly relevant information that we need to process all along our life. These meaningful sounds are crucial for social communication as well as for everyday life in our environment. Affective auditory information can be conveyed by many different sound sources spanning from human voices to traffic noises. The former being part of innate emotional communication while the latter is part of specific environmental sounds inducing emotional feelings specific to each person and its background. Processing of emotional sounds allow us to categorize everyday sounds to have an appropriate behavioural response, protect us from danger, and allow social interactions. If you hear the sound of an explosion, your brain will react by preparing you to run away, on the other hand, your own child laughing will induce in you a positive feeling. Besides the affective component, the localization of auditory objects is the key factor making sounds so influential in our life. In humans, the most developed and used sense is the vision, however, vision can be easily degraded or absent. Spatial hearing is thus very important to replace or complement vision to identify and correctly react to environmental events. Consequently, sounds, their emotional content and localization are significant information that our brain need to accurately process all along our life.

In this introduction, I will review first auditory principles such as the auditory cortex (AC) anatomy, auditory functions, auditory spatial representations as well as encoding of specific sound types. In a second part, I will introduce the concept of emotion in general and then more specifically emotional sounds. Finally, I conclude with some brief knowledge on fMRI principles.

## 1.2 Auditory system

### *1.2.1 Auditory pathways*

In brief, the primary ascending auditory pathway links the ear to the AC. It starts at the level of the cochlea, then reaches the cochlear nucleus located on the brainstem and then the superior olivary complex (SOC). Separate nuclei of SOC are responsible for the computation of spatial cues i.e., interaural time difference and interaural level difference (ITD and ILD respectively; see section 1.2.5 Auditory spatial representations). Neurons from SOC convey the auditory information to the midbrain in the inferior colliculus, and then the medial geniculate nucleus, located in the thalamus. From the thalamus, the neurons project finally to the AC (Brugge, 2013; Saenz and Langers, 2014).

### 1.2.2 Auditory cortex anatomy

Anatomical, histological and functional studies divided the AC into the core, belt and parabelt regions, with a clear hierarchical organization (Baumann et al., 2013; Clarke and Morosan, 2012; Rauschecker and Scott, 2009). Core and belt/parabelt areas can be differentiated on the basis of their cytoarchitecture. Studies on *post-mortem* brains allowed the identification of the Primary Auditory Cortex (PAC) based on architectonic landmarks (Morosan et al., 2001; Rivier and Clarke, 1997; Wallace et al., 2002). The previous studies described a high variability in PAC definition and localization, mainly pertaining to its major landmark, Heschl's gyrus (HG). HG presents a high intersubject and interhemispheric variability due to its different possible anatomies; single gyrus, partial or complete duplication (Da Costa et al., 2011; Moerel et al., 2014). This variability is even stronger in different populations with expertise in the auditory field, such as musicians and non-musicians (Benner et al., 2017). To unify the definition of PAC, a method has been developed, based on tonotopic mapping, meaning that neurons in the PAC are organized based on their tuning frequency (Da Costa et al., 2011; Formisano et al., 2003; Humphries et al., 2010; Striem-Amit et al., 2011; Talavage et al., 2004). PAC can be defined by the presence of two mirror-symmetric frequency progressions (high-low-low-high) running perpendicular to HG (Figure 1A and B). On the posterior part of HG, the first frequency gradient defines A1, while on the reverse direction, the region R can be delimited (Da Costa et al., 2011). Surrounding the two regions of PAC, four lateral regions can be characterized (L1, L2, L3 and L4), as well as four medial regions (M1, M2, M3 and M4; Da Costa et al., 2015, 2018; Figure 1C). These regions are less specific for frequency preferences than PAC and can be described as the belt and parabelt regions. This division of AC into ten regions will be used in the experiments described in this thesis.

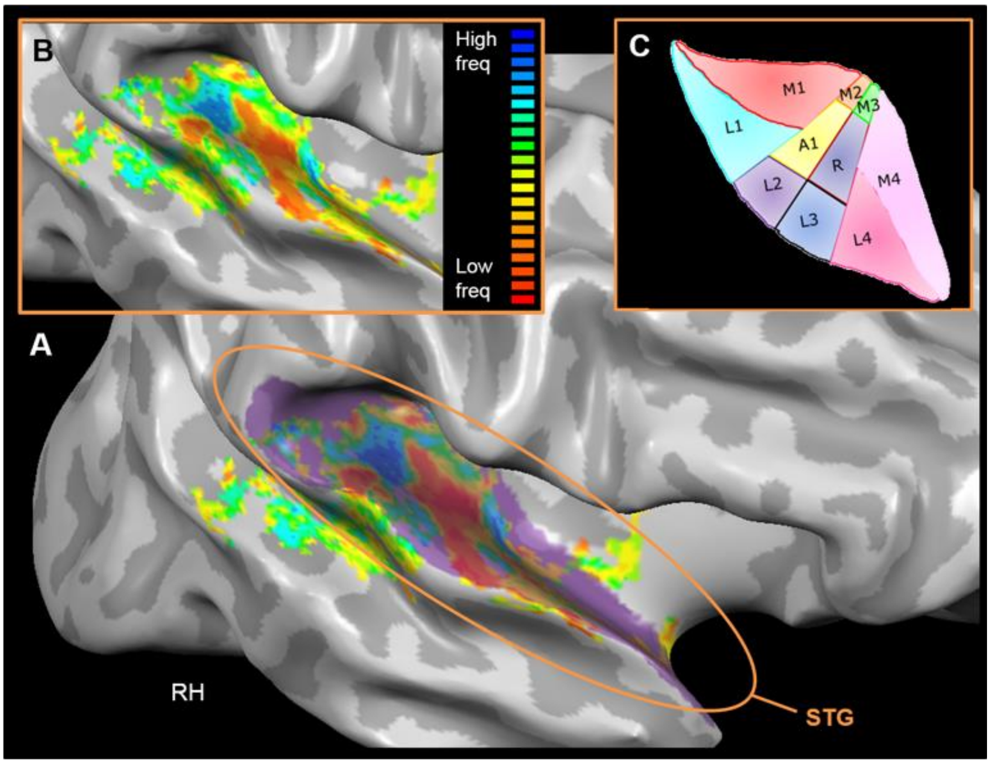

**Figure 1: Tonotopic mapping and auditory cortex division. A.** Tonopic map projected into the individual right hemisphere of a representative subject overlaid with the entire auditory cortex region (in purple). **B.** Tonotopic map projected into the individual right hemisphere of a representative subject. The warm colors represent the low frequencies while the cold colors represent the high frequencies. **C.** Division of the auditory cortex into 10 ROIs: A1 (yellow), R (brown), L1 (cyan), L2 (purple), L3 (blue), L4 (fuchsia), M1 (red), M2 (orange), M3 (green) and M4 (pink).

### *1.2.3 Auditory functions*

The hierarchical anatomical organization of AC coincides with a hierarchical processing going from pure tones to more complex auditory features and finally semantic aspects of auditory input (Binder et al., 2000; Chevillet et al., 2011; Cohen et al., 2016; Rauschecker and Scott, 2009; Rauschecker and Tian, 2000). PAC is involved in the processing of the most basic auditory objects, pure tones (Da Costa et al., 2011; Formisano et al., 2003; Talavage et al., 2004). Non-primary auditory areas are for their part tuned to more complex spectrotemporal features, such as frequency modulation, pitch, amplitude modulation or envelop (Altmann et al., 2008; Leaver and Rauschecker, 2016; Scott, 2005; Warren et al., 2005). These regions presented stronger response to band-pass noise or frequency-modulated sweeps compared with pure tones (Rauschecker and Tian, 2000; Wessinger et al., 2001). Higher-order areas respond to melody, linguistic and semantic aspects or specific sound categories

(Bergerbest et al., 2004; De Meo et al., 2015; Doehrmann et al., 2008; Hall et al., 2002; Kumar et al., 2007; Leaver and Rauschecker, 2010; Lewis et al., 2005, 2004; Lucia et al., 2010; Murray et al., 2008). The high anatomical variability found between different populations is also present at the functional level with bigger evoked responses for musicians than non-musicians (Schneider et al., 2002).

### *1.2.4 Auditory connectivity*

AC is a brain region displaying a high intra-auditory (between regions in AC) and extra-auditory areas (with other areas) connectivity. Within AC, studies described connections originating from PAC and reaching mainly the surrounding belt areas with only sparse direct connections between PAC and parabelt areas (Cammoun et al., 2015; Hackett and Kaas, 2004; LeDoux, 2000). The link to the parabelt areas is done by connections through the belt areas. Behond the belt areas, diffusion spectrum imaging (DSI) and functional connectivity studies described connections between the lateral belt areas and VA (Cammoun et al., 2015; Pernet et al., 2015). AC shows also strong connections with extra-auditory areas, such as medial geniculate body, frontal or parietal regions (Ethofer et al., 2012). Most of these studies described intra-hemispheric connectivity, but AC is also strongly connected at the interhemispheric level (Budinger and Heil, 2006; Kaas and Hackett, 2005).

### *1.2.5 Auditory spatial representations*

In our environment, the sounds are spatially positioned and our brain needs to determine the sound sources based on specific spatial cues in order to react properly. Sound localization in the horizontal plane is mostly supported by two mechanisms: interaural level difference (ILD) and interaural time difference (ITD) (Grothe et al., 2010). The difference either in time or in intensity allows the computation of the angle of the sound source in SOC. Sound localization is based on variable ratio of ITD and ILD in function of the acoustic features of the sound. ILD are more prominent in high frequency sounds, while ITD are the principal localization cues in low frequency sounds. ILD describe the fact that lateralized sound will have a different intensity for the left and right ear (Figure 2), while ITD reflect the fact that a lateralized sound will reach both ear at a different time, specifically a left-lateralized sound will first reach the left ear and then the right one.

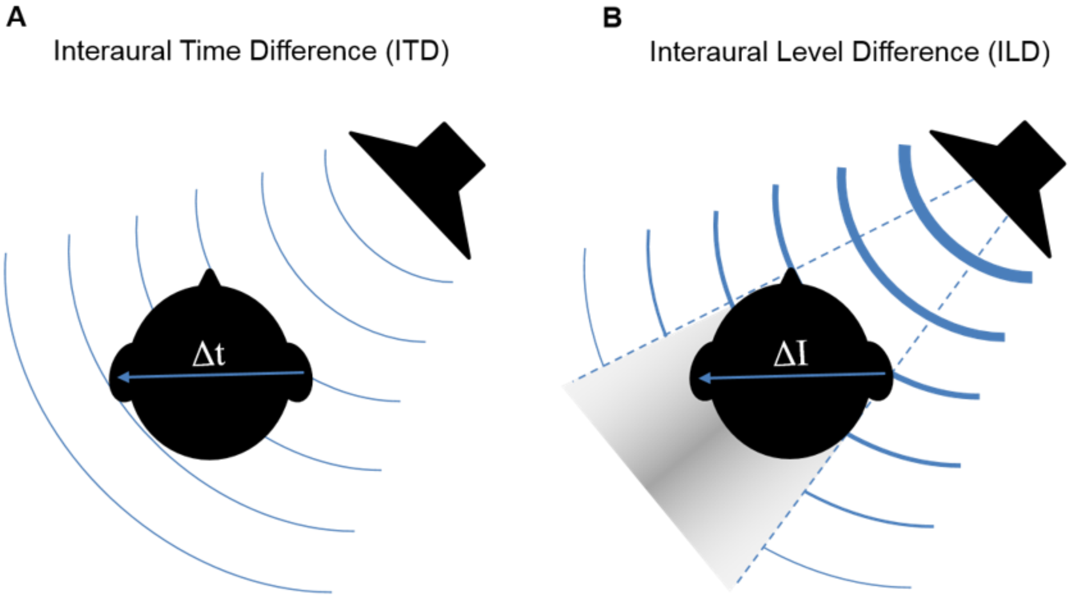

**Figure 2: Localization cues in the horizontal plane. A.** Interautal Time Difference (ITD). **B.** Interaural Level Difference (ILD). The thickness of the lines represent the intensity of the sound. The grey part represents the acoustic shadow produced by the head. Figure adapted from Grothe et al., (2010).

Higher in the processing pathway, a specific cortical network is responsible for the processing of auditory spatial information. This functional network is the dorsal stream, also called the "where" pathway, in opposition to the ventral stream, or "what" pathway, from the dual stream model (Rauschecker and Scott, 2009). This model is based on animal studies, human activation studies and brain lesion studies (Altmann et al., 2008; Clarke et al., 2000, 2002; Rauschecker and Tian, 2000; van der Zwaag et al., 2011; Viceic et al., 2006; Warren et al., 2002), and formulates that the ventral stream is selective for the meaning of the sounds, namely the recognition of categories of environmental sounds, as well as the processing of the semantic aspects of the auditory input (Altmann et al., 2007; Da Costa et al., 2015; Engel et al., 2009; Leaver and Rauschecker, 2010; Lewis et al., 2005; Murray et al., 2006). This "what" pathway runs anterior to PAC from the planum polare to the inferior frontal cortex. On the other hand, the dorsal stream is selective for the location of the sound source and runs posterior to PAC from the planum temporale and project to the parietal cortex. PAC is not part of the dual-stream model, as its specificity is equal for sound recognition or localization (Maeder et al., 2001). In parallel to this dual-stream model, evidence suggests that a third stream could be implicated in the processing of the integrated identity and position of a sound (Adriani et al., 2003; Altmann et al., 2007; Bourquin et al., 2013; Clarke and Geiser, 2015; Da Costa et al., 2018; van der Zwaag et al., 2011; Figure 3).

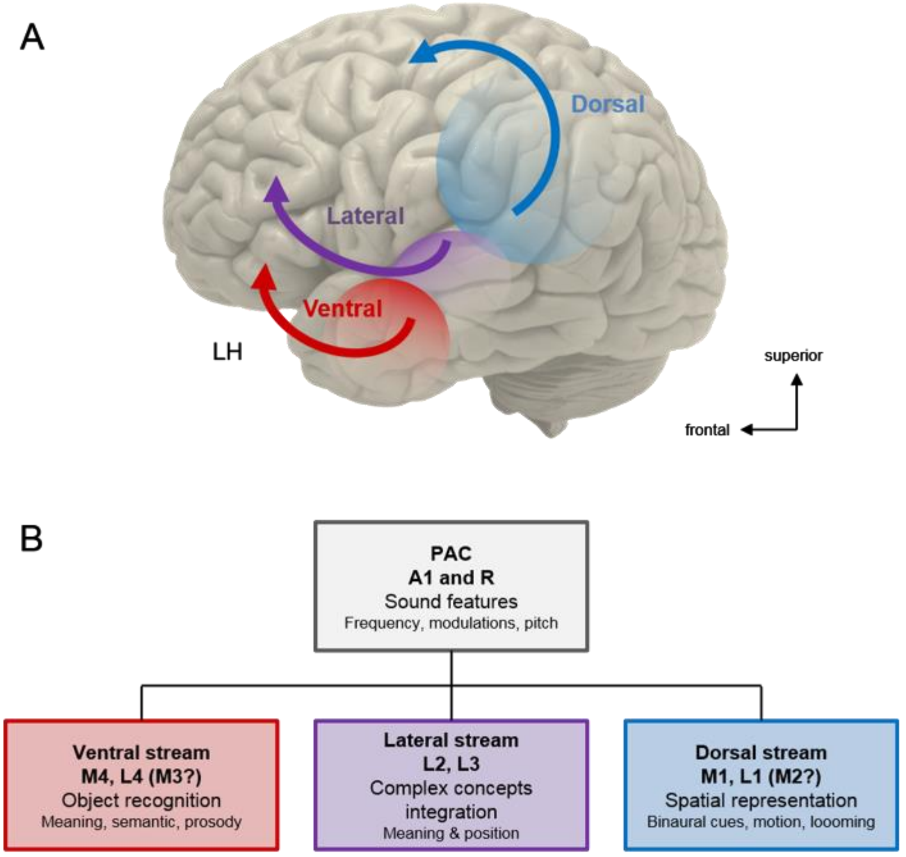

**Figure 3: Three-stream model of auditory processing. A.** Representation of three-stream model running from the auditory cortex to extra-auditory areas. The ventral «what» stream is represented in red, the dorsal «where» stream is in blue, while the lateral «integration» stream is depicted in purple. **B.** Repartition of the auditory cortex regions in the three-stream model and their principal characteristics. Figure adapted from Da Costa et al., (2018).

### *1.2.6 Voice processing*

In addition to this hierarchy, AC shows also fluctuating sensitivity to the harmonic content of the sounds. The regions the closer to PAC respond to harmonics present in artificial sounds, then more lateral regions (i.e., in the border of the superior temporal gyrus; STG) exhibited sensitivity to the harmonics of animal vocalizations, and finally even more lateral regions (i.e., in the border of the superior temporal sulcus; STS) respond to human vocalizations (HV), with the larger harmonic content (Brefczynski-Lewis and Lewis, 2017.; Giordano et al., 2014; Latinus et al., 2013). This latter region, namely the voice area (VA), show a high specificity for HV and is located in non-primary AC on the STS (Belin et al., 2004, 2002, 2000). These studies described VA as a region responding stronger to vocal sounds (speech and non-speech) than to any type of non-vocal sounds, comprising also scrambled speech sounds. In the right VA, the specificity for vocal sounds from human sources is even conserved against stimuli matched for low-level acoustic features, such as pitch or harmonics to noise ratio (Agus et al., 2017). This region is difficult to define anatomically, even if recent study described a coherence between the functional location of VA and the deepest point of the STS (Bodin et al., 2017). An overlap between the regions responding to low frequencies and voices was also described by Moerel et al., (2012). Moreover, individual functionally-defined VA are very variable in term of location and extent (Pernet et al., 2015). In addition to VA, auditory regions in the STG and non-auditory regions showed a preference for vocal sounds compared to non-vocal sounds, and are also part of a large voice network (Aglieri et al., 2018; Leech and Saygin, 2011). These regions included prefrontal regions (e.g., inferior prefrontal cortex) and subcortical structures, such as the amygdala (AMY) (Pernet et al., 2015). Finally, the processing of vocalizations by VA is modulated by the emotional value of the stimuli (Ethofer et al., 2012; Grandjean et al., 2005; Leitman et al., 2010). VA shows a stronger response to emotional voices, being positive or negative, in contrast with neutral voices. This preference for affectively-loaded stimuli was not limited to VA, but extended to PAC, non-primary auditory areas and frontal regions (Leitman et al., 2010).

## 1.3 Emotional processing pathways

### *1.3.1 Amygdala anatomy, function and connectivity*

AMY is a subcortical structure composed of several subnuclei showing different connectivity, architecture and function. Cytoarchitectonic studies divided AMY into three major groups of nuclei: laterobasal, centromedian and superficial (Figure 4; Amunts et al., 2005). The majority of the sensory inputs, including auditory, reach AMY through its laterobasal nuclei. The centromedial nucleus is a major output region for the emotional responses (Benarroch, 2014; LeDoux, 2007). Studies described a lateralized effect in AMY, with the right AMY showing a more rapid detection of the stimuli compared to the left AMY that could be involved in a more detailed evaluation of the stimuli (Sergerie et al., 2008; van der Zwaag et al., 2012; Wright et al., 2003). The exact function of AMY is still a matter of debate as it has been proposed to be a relevance detector (Sander et al., 2003), a social impact detector (Vrticka et al., 2013) or even a novelty detector (Pedersen et al., 2017). AMY shows strong connections with the sensory cortices, and in particular the auditory system. The auditory information reach AMY both through thalamic and cortical inputs. These connections are bidirectional as are the ones with the prefrontal cortex (Amaral et al., 1992; LeDoux et al., 1990; Woodson et al., 2000).

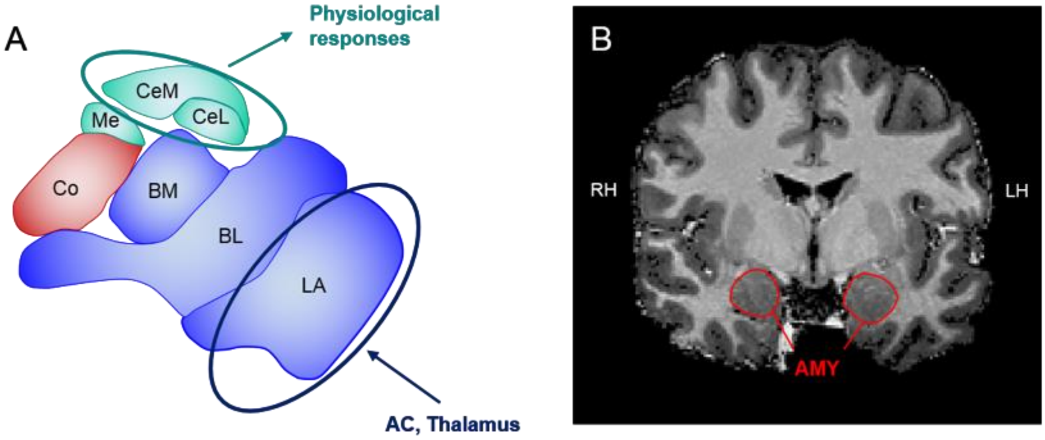

**Figure 4: Amydgala anatomy. A. Schematic representation of the amygdala nuclei.** The basolateral group is represented in blue, while the centromedian is in green and the superficial in red. LA: lateral nuclei, BL: basolatearal nuclei, BM: basomedial nuclei, CeM: centromedial nuclei, CeL: centrolateral nuclei, Me: medial nuclei, Co: Cortical nuclei. Figure adapted from Benarroch et al., (2014). **B.** Coronal view of a structural MRI with the amygdala encircled in red.

AMY is a region with a central role in emotional processing. However, its subcortical location in the brain contributes to the difficulty of imaging this small structure. Indeed it is located in the vicinity of large veins (Boubela et al., 2015), in a region with inhomogeneity in the local magnetic field (Labar et al., 2001) and subject to a strong dephasing (Mathiak et al., 2012), which induce an increased susceptibility to artefacts and a reduced signal-to-noise ratio (SNR) from AMY. Studies proposed specific functional magnetic resonance imaging (fMRI) acquisition parameters to improve the imaging of AMY (van der Zwaag et al., 2012), however these are difficulty reconcilable with the parameters needed to image the subregions of AC, such as a slab of limited thickness with a specific angle.

### *1.3.2 Emotion processing*

Emotional processing engages a large network of multiple regions (Dalgleish, 2004; Duerden et al., 2013; Lindquist et al., 2012; Peelen et al., 2010; Phan et al., 2002). These regions comprise the insula, the prefrontal cortex, AMY, hippocampus, orbitofrontal cortex and anterior cingulate cortex. Their involvement was defined with lesions or activation studies. The implication of AMY in the processing of emotions was first established for negatively valenced stimuli (Morris et al., 1998; Phillips et al., 1998; Wright et al., 2001), but is now well documented for positive stimuli (Anderson et al., 2003; Ball et al., 2007; Costa et al., 2010; Hurlemann et al., 2008; O'Doherty et al., 2001; Sergerie et al., 2008; Winston et al., 2005; Zald, 2003). This valenced-view of the emotions is challenged by hypotheses concerning the evaluation of the arousal of the stimuli and not its valence by AMY (Anderson et al., 2003; Bonnet et al., 2015; Viinikainen et al., 2012; Zald, 2003). Regardless that AMY encodes the valence or the arousal of the stimuli, it processes behaviourally relevant stimuli with a very quick processing of the emotional value (Sauter and Eimer, 2009). Moreover, emotions modulate many physiological variables. This is the case of the heart rate, heart rate variability, pupil size dilation, corrugator activity (facial electromyography), startle reflexes and skin conductance (Bradley and Lang, 2000; Brouwer et al., 2013). Mainly, the heart rate shows a strong deceleration in response to negative stimulation (Bradley et al., 2008; Gomez et al., 2005; Martin-Soelch et al., 2006) and the pupil size dilation increases significantly in response to emotional compared to neutral stimuli (Bradley and Lang, 2000; Partala and Surakka, 2003). In emotional experiments, it is important to record physiological variables and to use them either as regressors in the analyses to remove physiological noise from the data or as implicit measure of valence. Brain activation and physiological variables modulation are correlated with several personality traits, mostly anxiety and depression (Canli et al., 2001; Frühholz et al., 2017; Laeger et al., 2012). For instance, high anxiety scores are associated with increased heart rate deceleration (Martin-Soelch et al., 2006) as well as increased AMY activation (Stein et al., 2007).

### *1.3.3 Emotional sounds processing*

In the auditory domain, a large network was described to be responsible for the processing of the emotional content of sounds (Figure 5; Frühholz et al., 2016; Schirmer and Kotz, 2006). AMY, VA and early-stage auditory areas are part of this system. It is a core network composed of cortical and subcortical regions that processes emotional sounds in general. For the processing of specific categories of affective auditory stimuli specialized regions (e.g., orbitofrontal cortex, thalamus, inferior colliculi, hippocampus) are engaged in complement to the core network.

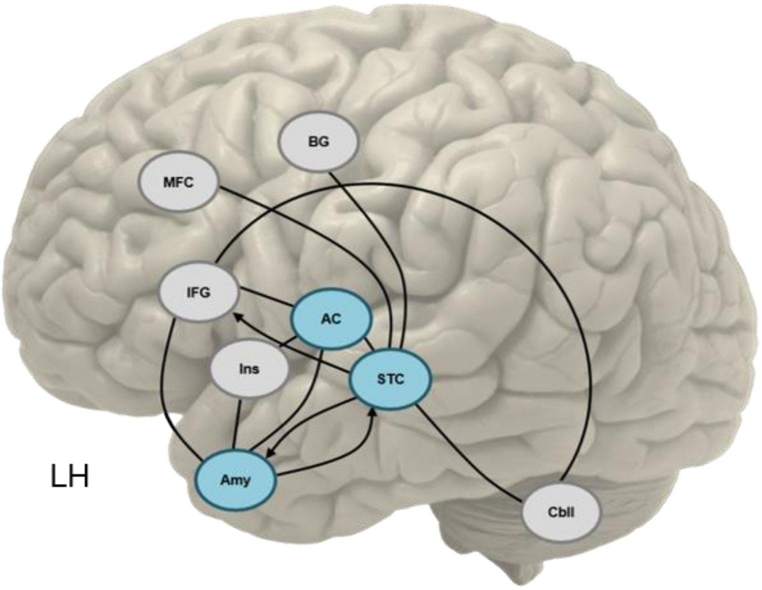

**Figure 5: Core network of emotional sounds processing.** Lines represent functional connections between two regions, while arrows depicted unidirectional connections. The regions in the scope of the current thesis are highlighted in blue. MFC: medial frontal cortex. IFG: inferior frontal gyrus. Ins: Insula. AC: auditory cortex. STC: superior temporal cortex. BG: basal ganglia. Amy: amygdala. CBll: cerebellum. Figure adaptated from Frühholz et al., (2016).

The modulation of auditory processing by emotional valence has been described with HV (Fecteau et al., 2007; Morris et al., 1999; Pell et al., 2015; Phillips et al., 1998; Sander and Scheich, 2005, 2001), prosody (Ethofer et al., 2008; Frühholz and Grandjean, 2013; Johnstone et al., 2006; Leitman et al., 2010; Wiethoff et al., 2009), music (Aubé et al., 2015; Koelsch, 2010) and environmental sounds (Plichta et al., 2011; Viinikainen et al., 2012). In their study, Fecteau et al., 2007 described an increased activity of AMY, PAC and non-primary auditory areas in response to emotional vocalizations compared to neutral ones. This preference of AC for affective content was also present when tested with emotional environmental sounds (Plichta et al., 2011).

## 1.4 Methodology

### *1.4.1 Sound battery*

In the experiments of this thesis, we used a sound battery tested in a behavioural paradigm (Aeschlimann et al., 2008) showing that the HV are a special category of sounds compared to environmental sounds. This battery is composed of 66 emotional sounds of 2 seconds, distributed into six categories: Human Vocalizations Positive (HVP), Human Vocalizations Neutral (HV0), Human Vocalizations Negative (HVN), Non-Vocalizations Positive (NVP), Non-Vocalizations Neutral (NV0), Non-Vocalizations Negative (NVN). The HVN category is composed of sounds of screams and fighting, the HV0 of syllables spoken either by a man or by a woman, while the HVP are laughs or erotic sounds. The non-vocalization categories are composed of environmental sounds from diverse sources; here are some example for each categories. NVN: gunshot, alarm clock and thunder; NV0: wind, train and court sport; NVP: applause, beer and guitar. The acoustic characteristics of the different sound categories are described in Grisendi et al., 2019a.

### *1.4.2 Ultra-high field imaging*

The investigation of brain activation at 3 Tesla is limited by its spatial resolution. Using ultra-high field (≥7 Tesla (7T)) scanner allows the study of small structures, such as AMY (Sladky et al., 2013), or specific small ROI, such as subregions of AC (Da Costa et al., 2015, 2018), thanks to its increased sensitivity and specificity (De Martino et al., 2017; Dumoulin et al., 2017). The major advantage of imaging at 7T is the increased SNR which increments linearly with the magnetic field strength (Talavage et al., 2014; Triantafyllou et al., 2005; van der Zwaag et al., 2009, 2011). The increased SNR has three benefits: the use of smaller voxel size, the limited use of spatial smoothing, and the feasibility of single-subject analysis. However, the increase in SNR is counterbalanced by an increase in magnetic field inhomogeneity and physiological noise. This increase in artefacts make the imaging of structures located near air/water interfaces even more difficult. The imaging can be improved with specific acquisition parameters (van der Zwaag et al., 2012), as well as with the use of the physiological variables as regressors (Kasper et al., 2017; Reynaud et al., 2017).

### *1.4.3 BOLD signal and fMRI designs*

fMRI is a non-invasive indirect method used to measure brain activity (Logothetis et al., 2001; Poldrack et al., 2011; Talavage et al., 2014). This method infers brain metabolic activity based on measures of the changes of blood oxygenation, namely on the Blood Oxygenation Level Dependent (BOLD) signal. This measure is based on the evidence that the haemoglobin in the brain display different magnetic properties depending on its state as oxygenated (Hb) or deoxygenated (dHb). Specifically, dHb is paramagnetic, while Hb is resistant to the magnetic field. An increase in the ratio of dHb/Hb induces a reduction in the local magnetic field homogeneity and thus a diminution of the signal. When a region of the brain is activated, its neurons require more oxygen, implicating an increased blood flow and blood volume and thus a decrease of the ratio dHb/Hb leading to a signal increase. The hemodynamic response reflects the modulation of the blood flow required by the activation of neurons and then their deactivation.

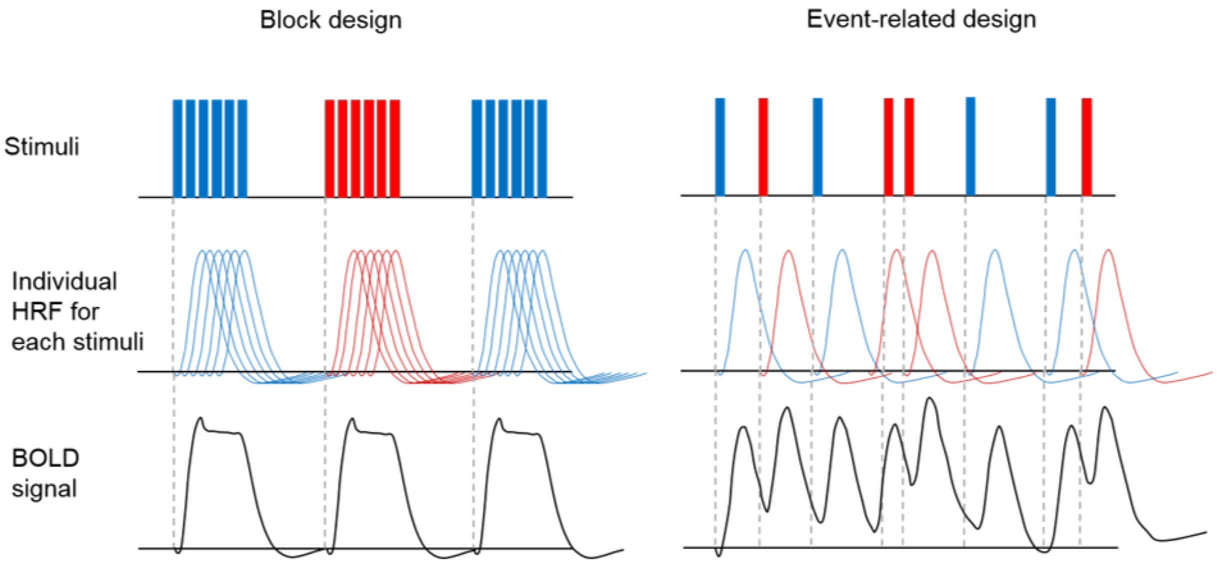

**Figure 6: Block *vs.* Event-related design.** Representation of an experiment with a block design (left panel) or an event-related design (right panel). The first line represents the experimental paradigm, with two different types of stimuli depicted in blue or in red. The second line illustrates the individual hemodynamic response function (HRF) induced by each separate stimuli. The last row represents the estimated BOLD signal (summation of the individual HRF). This simplified representation is valid only in case of imaging of a region with neurons with the same sensibility for both stimuli.

fMRI experiments are constructed either with block designs or event-related designs (Figure 6). Event-related designs are based on the discrete presentation of stimuli that are separated by inter-stimulus interval of rest. This type of experimental design increases the design flexibility and permit the investigation of individual trial responses, as the hemodynamic response is estimated for each event. The major drawbacks of using event-related designs are the low SNR, the duration of the experiments and the complexity of the analyses in function of the inter-stimulus time (in case of overlapping of the hemodynamic responses, the signal need to be deconvolved). In order to strengthen the signal, the number of trials needs to be increased which results in longer scanning time. In the other hand, block designs are built on on/off patterns, with periods (blocks) of stimuli presentation interleaved with periods of rest. Block designs are advantageous in term of design and analysis simplicity and mostly of its strong statistical power. This is counterbalanced by the loss of the information for individual response time-course. The presentation of the stimuli in blocks induce a summation of the hemodynamic response, thus increasing the amplitude of the response but losing the information about individual activation. Block design allow to study the specific effect of repetition suppression that is a decrease in neural activity following the repeated exposure to the same feature (Barron et al., 2016; Grill-Spector et al., 2006). This effect is used to study the sensitivity and specificity of brain regions to a given feature of stimuli or to a given category of stimuli. It allows also for the disentangling if some regions are made of different populations of neurons responding to different characteristics.

## 1.5 Aim of the thesis

In this thesis, I studied the encoding of emotional sounds in a non-spatial or spatial context in AC, VA and AMY and addressed three main issues:

1. The relative contribution of early-stage auditory areas, VA and AMY to the processing of emotion and vocalization and to their integration.
2. The modulation by the spatial origin of the encoding of emotional human vocalizations and emotional environmental sounds in early-stage auditory areas and VA.
3. The modulation of the functional correlations between early-stage auditory areas and VA by sound category, valence and lateralization.

# 2 Summary of the results

## 2.1 Study A: Processing pathways for emotional vocalizations

Tiffany Grisendi, Olivier Reynaud, Stephanie Clarke and Sandra Da Costa

This article is accepted in the journal “Brain Structure and Function”.

*Contribution*

The candidate contributed in the experimental design, recruitment of the participants, data acquisition, data analyses and manuscript preparation

*Abstract*

Emotional sounds are processed within a large cortico-subcortical network, of which the auditory cortex, the voice area and the amygdala are the core regions. Using 7T fMRI we have compared the effect of emotional valence (positive, neutral, negative) and the effect of the type of environmental sounds (human vocalizations, non-vocalizations) on neural activity within individual early-stage auditory areas, the voice area and the amygdala. A 2-way ANOVA was applied to the BOLD time course within each ROI. In several early-stage auditory areas it yielded a significant main effect of vocalizations and of valence, but not a significant interaction. Significant interaction as well as significant main effects of vocalization and of valence were present in the voice area; the former was driven by a significant emotional modulation of vocalizations but not of other sounds. Within the amygdala only the main effect of valence was significant. Post hoc correlation analysis highlighted coupling between the voice area and early-stage auditory areas during the presentation of any vocalizations, and between the voice area and the right amygdala during positive vocalizations. Thus, the voice area is selectively devoted to the encoding of the emotional valence of vocalizations; it shares with several early-stage auditory areas encoding characteristics for vocalizations and with the amygdala for the emotional modulation of vocalizations. These results are indicative of a dual pathway, whereby the emotional modulation of vocalizations within the voice area integrates the input from the lateral early-stage auditory areas and from the amygdala.

## 2.2 Study B: Emotional value of the auditory space

Tiffany Grisendi, Stephanie Clarke and Sandra Da Costa

This article is in preparation for submission

*Contribution*

The candidate contributed in the experimental design, recruitment of the participants, data acquisition, data analyses and manuscript preparation

*Abstract*

Evidence from behavioural studies suggests that the spatial origin of sounds may influence the perception of emotional valence. Using 7T fMRI we have investigated the impact of the type of sound (vocalizations; non-vocalizations), emotional valence (positive, neutral, negative) and spatial origin (left, centre, right) on the encoding in early-stage auditory areas and in the voice area. The combination of these different characteristics resulted in a total of 18 conditions (2 Types x 3 Valences x 3 Lateralizations), which were presented in a pseudo-randomized order in blocks of eleven different sounds (of the same condition) in 12 distinct runs of 6min. In addition, the subjects (N = 14, with normal hearing) also listened to two different localizers (a tonotopy paradigm and a voice localizer), which were used to define the regions of interest. A 3-way repeated measure ANOVA on the BOLD responses revealed bilateral significant effects and interactions in the primary auditory cortex, the lateral early-stage auditory areas, and the voice area. Positive vocalizations presented on the left side yielded greater activity in bilateral primary auditory cortex than did neutral or negative vocalizations or any other stimuli at any of the three positions. The voice area did not share the same preference for the left space; spatial attributes modulated its activation by sound objects conveying positive or neutral emotional valence when presented on the right or left side (but not at the centre). Comparison with a previous study indicates that spatial cues may render emotional valence more salient within the early-stage auditory areas.

## 2.3 Study C: Functional correlations between early-stage auditory areas and the voice area

Tiffany Grisendi, Stephanie Clarke and Sandra Da Costa

This article is in preparation for submission

*Contribution*

The candidate contributed in the experimental design, recruitment of the participants, data acquisition, data analyses and manuscript preparation

*Abstract*

Human vocalizations processing and their modulation by emotional valence and/or localization involves brain regions, such as voice area (VA) and early-stage auditory areas. Using two separate datasets acquired at 7T fMRI, we have investigated the functional correlations between early-stage auditory areas and VA as modulated by the category of sound, valence and position. The functional correlations between VA, primary auditory cortex (PAC) and lateral belt areas were strongly modulated by the category of sound, weakly by the spatial positioning and not by the affective content. Human vocalizations produce stronger functional correlations between VA, PAC and lateral belt areas, compared to non-vocal environmental sounds.

# 3 Discussion

## 3.1 General discussion

The aim of this thesis was to investigate the processing of HV and NV and their modulation by emotional and/or spatial content within the early-stage auditory areas, VA and AMY. We acquired data of two block-design fMRI studies at 7T, using an emotional sound battery (Aeschlimann et al., 2008). The BOLD time-courses of the different sound categories were analysed with repeated measure ANOVAs. To complete this analysis we looked also at the functional correlations between our regions of interest (ROI) and how these correlations were modulated by the different characteristics of the sounds. In this chapter, I will first review the results of the first two studies and the impact of using lateralized sounds. Then I will discuss the model derived from the last study in light of current proposed model of emotional processing. Finally, I will focus on future directions using connectivity analyses as well as the interest of the results of this thesis for specific populations of patients.

### *3.1.1 Brief discussion of the results*

The different studies described in this thesis demonstrated that the emotional content modulate the processing of sounds in early-stage auditory areas, regardless of the sound category. In contrast, the emotional modulation is restricted to HV when processed in VA. This result was established using non-spatialized as well as lateralized sounds.

With non-lateralized sounds, our study (Grisendi et al., 2019a) highlighted that the stimulus category does not influence the emotional processing in AMY. The processing of emotions by AMY was already demonstrated by various studies with HV (Fecteau et al., 2007; Frühholz and Grandjean, 2013; Phillips et al., 1998) and with environmental sounds (Viinikainen et al., 2012; Zald and Pardo, 2002). However, most of the studies did not included both categories of sounds and thus did not compare the emotional encoding for HV or NV in AMY.

Specifically for lateralized sounds, our study (Grisendi et al., 2019b) demonstrated that there was a modulation of the HV by the emotional valence but not by the spatial origin of the sounds, in VA. The processing of positive sounds when presented in the left space, by PAC bilaterally, exhibited a strong sensitivity to the category of sound, with stronger response to positive HV than positive NV. This preference for HV was not present for any other sound position or emotional valence. Thus, the left auditory space appears to favour the specific encoding of positive HV within PAC in both hemispheres. In line with our results, Kryklywy et al., (2013) reported an enhanced activity of AC in the right hemisphere in response to positive stimuli located in the left hemispace compared to positive stimuli in the right hemispace.

As we used lateralized sounds of different categories, our results gave us an insight into the dual/three-stream model of auditory processing. However, our experiments were not designed to study specific effects in these processing pathways. Future investigations could study the impact of emotional processing on the dual/three-stream pathways. To do so, we should implement a "what" and a "where" pathway localizers that are distinct from the main experiment, as we did for the VA, with the voice localizer, and for AC, with the tonotopic mapping paradigm. These localizers would allow to define which specific regions are part of which auditory streams, and which regions are common to both pathway and are thus part of the third integration stream. Based on this repartition of our ROIs we could investigate the emotional modulation in each stream and highlight if there is a special processing within these auditory streams for the socially relevant stimuli. Kryklywy et al., (2013, 2018) documented an emotional modulation of the "what" auditory pathway that was not paralleled by an emotional modulation of the "where" stream. Moreover, they described an anterior-lateral region of the right AC that could be part of the third-stream for the specific processing of emotional sounds. This region could correspond to our right L3 ROI that showed a preference for the processing of positive HV in the lateral space.

In the third study, we compared Study A and B (Grisendi et al., 2019a,b) to study the impact of using lateralized sounds in an experiment on the processing of HV and other environmental sounds and their modulation by emotional content. However, the next discussed results are only exploratory and descriptive as the setup of both experiments was not identical and the difference was not limited to the use of lateralized sounds or not. More specifically, the second experiment present more subjects, a coil with more channels as well as a different way to define VA. These changes could have an influence on the statistical power or on the activation extent. But part of these descriptive results can be account for the use of lateralized sounds. Investigating the impact of using lateralized sounds highlighted that the processing of vocalizations is not modulated by the presence of spatial cues, as the same set of areas process preferentially HV in both studies. On the other hand, the processing of emotions seem to be modulated by the presence of spatial cues, as more regions are implicated in this processing in Grisendi et al., 2019b. The interaction Valence x Lateralization revealed only a significant effect in VA and not in any early-stage auditory areas, indicating that this modulation is specific to the spatial context of the experiment and not to the sound position itself. This means that the presence of spatial cues increases the saliency of sounds with an affective content. This could be due to a more realistic experimental setup in case of lateralized sounds, as the sounds in the everyday life are always in a spatialized environment, which could lead to stronger emotional response or enhanced discrimination of emotional valence.

Based on results of Grisendi et al., 2019a, (with non-lateralized sounds), we described a model of dual input to VA with category-specific input from lateral belt areas and emotion-specific input from AMY. This model was only based on functional correlations between BOLD signal of the ROIs and did not give us any insight about the direction of the connections. Different theoretical models were proposed for the processing of emotional sounds but there is currently no consensus (Kumar et al., 2012; Liebenthal et al., 2016; Pannese et al., 2015; Schirmer and Gunter, 2017; Tschacher et al., 2010). For instance, using dynamic causal modelling (DCM), Kumar et al., (2012) described a model with bidirectional connections between AC and AMY, which are modulated by different features (i.e., acoustic features and valence). For their part, Pannese et al., (2015) proposed a model with multi-step processing, including an early decoding based on acoustic features at a subcortical level, associated with a higher-level processing at the cortical level. With a simplified model of both studies (Figure 7; correlations with adjusted $R^2 > 0.9$), we can observe that the lateral regions of the auditory belt are more implicated in the processing of emotional sounds than the medial regions, especially with non-lateralized sounds. However, due to our focus on auditory regions and the type of analyses we did, we cannot favour one model more than another one. Future connectivity studies (described in 3.3.1 Connectivity) could give answers to this question.

In both studies we described a modulation of the functional correlations between VA and auditory areas in function of the category of sound and not related to the emotional content of the sound. The processing of HV, and not of environmental sounds, implicate strong functional correlations between VA, PAC and lateral belt areas, whatever the position of the sound or its valence. This model is based only on ITD spatial cues, and future studies with ILD or virtual reality using a combination of ITD and ILD, would be necessary to generalize this model to all types of lateralized sounds.

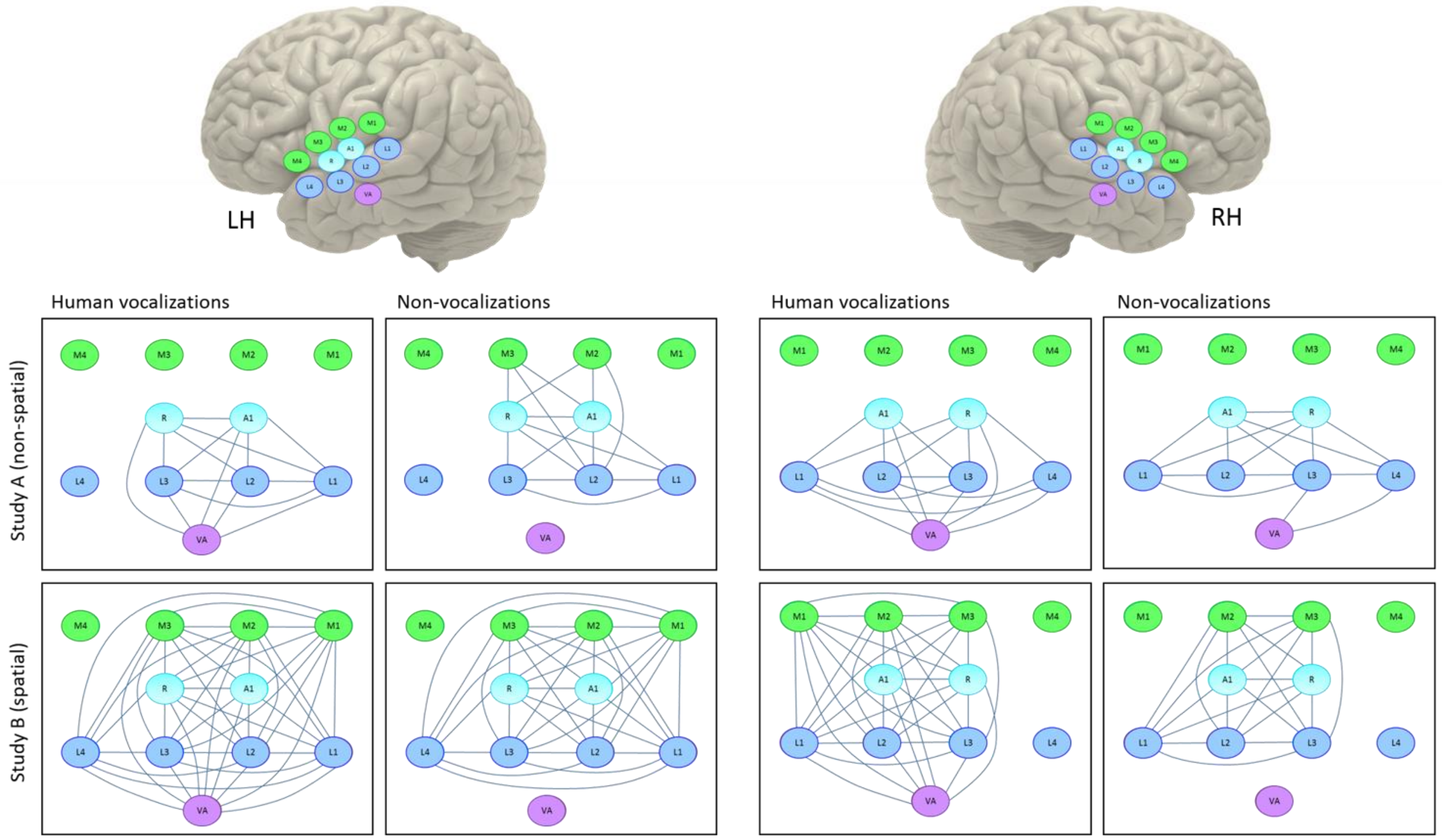

**Figure 7: Model for emotional sounds processing within the human auditory cortex.** Correlations between the ROIs of the medial belt area (in green; M1, M2, M3 and M4), PAC (in cyan; A1 and R), the lateral belt area (in blue; L1, L2, L3 and L4) and VA (in purple), in the left and right hemispheres separately (left and right panel respectively). The correlations are separated for the human-vocalizations and non-vocalizations. The first line represents the correlations of the BOLD signals from the study A (Grisendi et al., 2019a) with non-lateralized sounds, while the second line depicted the results from the study B (Grisendi et al., 2019b) with lateralized sounds. For a simplification of the representation, correlations with an adjusted- $R^2$ smaller than 0.9 are not represented. The three different valences (positive, neutral, negative) and the three lateralizations (left, center, right) are grouped together. LH: left hemisphere. RH: right hemisphere.

In study C (Grisendi et al., 2019c), we demonstrated a dissociation of VA correlations in function of the vocalization content. A follow-up study would include functional connectivity with different sound categories to confirm the present results. Moreover, our analyses were restricted to AC, and did not show a global picture of the processing. Two main reasons rise the importance of integrating extra-auditory regions in the analysis. First, there is strong evidence of a core network implicated in the processing of emotional sounds composed of auditory and non-auditory regions: medial frontal cortex, inferior frontal gyrus, insula, AC, basal ganglia, AMY and cerebellum (Frühholz et al., 2016). Secondly there are proofs for connections between AC and extra-auditory regions (Budinger and Heil, 2006; Kaas and Hackett, 2005). Finally, the involvement of different population of subjects with different auditory expertise would be interesting to study the modulation of the functional connectivity in these specific situations, as auditory anatomy and functions are known to vary across different populations.

### *3.1.2 Limitations and improvements*

The sound battery used in our experiments was composed of emotional sounds, varying in term of valence (positive, negative or neutral) and type (HV, NV). The different categories were controlled for their general acoustic characteristics, more specifically their mean spectrograms were not statistically different. However, this global control did not guarantee that the categories are not acoustically different along specific parameters. We found mainly big differences for the category HV0, which was statistically different from all other sound categories in term of entropy and harmonics-to-noise ratio. The type of sounds present in each category can account for this. Indeed, the HV0 category was composed exclusively of speech sounds, while HVN and HVP categories did not contain any speech sounds, and were composed of non-verbal vocalizations only. HV, specifically speech sounds, contain greater harmonics-to-noise ratio than other environmental sounds (Lewis et al., 2005), and this could be implicated in the specific processing of voices. Moreover, the BOLD signals of VA indicate that the sounds from the HVN category are not recognized as pure vocalizations as the response is lower than for HVP or HV0 categories. This finding was present in both experiments, and is thus due to the sound battery. Finally there are repetitions of different exemplars of the same sound objects within a category (mostly NVP, HVN and HVP) as well as between different categories. This could be a problem as evidence showed that a repetition priming effect is induced by the presentation of sound objects from the same sound source (Bourquin et al., 2013; Da Costa et al., 2015). All the previous points could lead to confound factors in our experiments and thus emphasize the importance of using a controlled sound battery. However, it is impossible to control for every acoustical parameters without ending with categories composed of identical sounds (Leaver and Rauschecker, 2010). This is especially true when using emotional sounds, as studies documented a discrimination of valence of vocalizations by AC based on acoustic features (Bestelmeyer et al., 2017; Ethofer et al., 2006; Leitman et al., 2010).

The repartition of the sounds in a given category of valence was based on a separated behavioural study with a different population of subjects (Aeschlimann et al., 2008). The perception of the sound valence is linked to each individual subject and relates to its own life experiences. An implicit rating of the sounds valence would be beneficial to study the processing of different emotions without a-priori categorization. This could be achieved with the help of eye-tracking. This method measures the amount of pupil size dilation and reflects the emotional load of the stimuli. The pupil dilation increases for emotional stimuli compared to neutral stimuli, and this is the case for positive and negative valences (Bradley et al., 2008; Partala and Surakka, 2003). By using this method of implicit valence judgment we could make a personalized repartition of the sounds in the emotional categories and thus have a perfect match between the valence perceived by the subject and the valence category in which the sound is assigned for the statistical analysis. This would increase the sensitivity of the analysis. Moreover, implicit or explicit (as valence ratings in Grisendi et al., 2019a) valence judgment could be used to re-analyse the data based on an event-related point of view.

In the study investigating the representation of emotional sounds (Grisendi et al., 2019a), the quality of the structural imaging did not allow us to anatomically parcellate AMY. This would be of special interest, as the different subnuclei of AMY are known to present different functional responses and connectivity patterns (Benarroch, 2014; Bzdok et al., 2013; Kim et al., 2011; LeDoux, 2007; Sah et al., 2003). The use of the whole AMY as a ROI could impair the visualisation of results with an averaging of the responses from the different subnuclei. For instance, Ball et al., (2007) reported a positive signal change in the basolateral group at the same time as a negative signal change in the superficial and centromedial groups in response to emotionally-loaded music. With positive and negative signal changes in response to the same stimuli within AMY, the global activation would show no or only little response. This highlighted the importance of a parcellation of AMY in studies investigating the processing of emotions, and thus the use of high spatial resolution and specific structural MRI sequences.

In the study investigating the spatial representation of emotional sounds (Grisendi et al., 2019b), AMY was not included in the analysis, as this region did not show any response in any of the experimental conditions. This was a major drawback as AMY is known to be implicated in the global processing of emotional sounds (Frühholz et al., 2016) and was responsible of the analysis of the affective content of the sounds from our battery as shown in the first study (Grisendi et al., 2019a). We hypothesize that this lack of results could be due to the type of spatial cues used to lateralize the sounds (ITD). Even if both ITD and ILD are first compute in the SOC, they are processed in two different nuclei along different pathways (Grothe et al., 2010; Tardif et al., 2006). ILD are first processed in the lateral superior olive (LSO) while ITD are initially computed in the medial superior olive (MSO). These different pathways could explain an insensibility of AMY to the ITDs. To test for this hypothesis we should reproduce the same study with sounds lateralized by ILD spatial cues. The analyses of Grisendi et al., (2019b) were thus limited to AC. Moreover, in this study, the use of lateralized sounds did not allow us to reveal regions from the “where” pathway, as only regions of PAC in the left hemisphere reported to show a main effect of lateralization. This could be due to the lack of localization task, or change of sound position in the block, in our experimental paradigm.

The impact of using lateralized sounds on emotional sounds processing could be studied by comparing the same experiment using once lateralized sounds, and once non-lateralized sounds. However, in this thesis, this was limited by divergence between both studies. To study only the impact of spatial context and to avoid any other confound factors, we would need to have a perfect match between the setup of both studies, namely the same subjects (pseudo-randomly assigned to participate in study A or B first), the same MRI material (scanner and coil), the same fMRI paradigm (number of sound presentations, duration of the blocks, instructions) and the same ROIs definition (localizer). This was not the case of our studies, which restricted our conclusions about the use of lateralized sounds. To have an insight into the differences between both studies, we performed a 3-way ANOVA Vocalization (HV, NV) x Valence (negative, neutral, positive) x Study (study A, study B) on the data of four subjects who participated in both studies. This statistical analysis did not reveal any effect of the factor Study, suggesting thus a weak influence of the varying parameters between both studies.

In both studies, our analyses were restricted to specific ROIs because of the high spatial resolution required to image individual sub-regions of AC, and thus the limited field of view we could use. This assumed choice was a big advantage to investigate the specific processing of emotional sounds within individual sub-regions of AC, but did not allow us to have a global picture of the whole network. In future studies investigating the entire emotional sounds network, we should use a whole brain acquisition (Narsude et al., 2016; van der Zwaag et al., 2018).

## 3.2 Conclusions

Emotional sounds are highly salient auditory stimuli that we need to process adequately in our everyday life in order to protect us from danger and to promote social interactions. The localization of this affective information is important to enable an accurate behavioural response. The aim of this thesis was to investigate the encoding of sounds with variable vocalization, emotional and spatial contents within the early-stage auditory areas, VA and AMY.

Our results demonstrated an encoding of the valence of the stimuli in the early-stage auditory areas and AMY independently of the category of sounds. The emotional processing in VA was for its part limited to the category of HV. A category-specific coupling between VA and early-stage auditory areas paralleled this restricted emotional modulation. We found strong functional correlations between VA and auditory lateral belt areas for the presentation of HV lateralized or non-lateralized. The presence of spatial cues modulates only the processing of valence and not of vocalization, and increases the functional coupling of PAC and auditory medial belt areas with aforementioned regions. Finally, we described a favoured encoding by bilateral PAC of the positive HV when presented in the left auditory space.

These findings give us new insights into the processing of affective sounds by early-stage auditory areas, VA and AMY and strengthen the importance of VA as a computational hub for the processing of emotional vocalizations.

## 3.3 Future perspectives

### *3.3.1 Connectivity*

In the last part of this thesis we described the functional correlations between VA and early-stage auditory areas. It would be useful to design a functional connectivity study to investigate the connections between our ROIs but also with other areas involved in the processing of emotional sounds. Several studies showed increased functional coupling between cortical voice areas and subcortical structures such as basal ganglia and thalamus, when processing emotional prosody (Ethofer et al., 2012; Frühholz and Grandjean, 2012). Moreover, Koelsch et al., (2018) described different functional connectivity of AC when processing affective music. Their study highlighted an intra-auditory network, with functional connections between PAC and other auditory regions (i.e. contralateral PAC and non-primary auditory areas), and an extra-auditory network, with functional connections between anterior and posterior regions of auditory association cortex with limbic, somatosensory, visual and attentional structures. However, none of these studies documented global functional connectivity of subregions of AC with other auditory areas and extra-auditory areas.

Future connectivity studies could investigate different hypotheses:

1. Seed-based functional connectivity with ROIs from localizers (AC, VA, "what" and "where" regions) would highlight the connections as well as the segregation and integration between the auditory streams in case of emotional processing. This seed-based functional connectivity experiment could investigate the connections of the different seeds to specific structures, such as AMY, and study if they target the same or different part of this region.
2. Functional connectivity with a whole brain approach could investigate the global network implicated in the processing of emotional sounds. This would give an insight into the whole emotional sounds network, as the majority of studies focused only in parts of this network.
3. Eigenvector centrality mapping (Lohmann et al., 2010) could be used to investigate the connections within the emotional network and define computational hubs. This method assigns a value to each voxel in function of the number of connections of the given voxel and the centrality of the voxels to which it is connected. Big value corresponds to a voxel strongly connected with central nodes of the network. We could investigate if the computational hubs of emotional sound network are modulated by the valence, the category or position of the stimuli.
4. Vector autoregression (Roebroeck et al., 2005) would take advantage of the BOLD time-series to investigate causal relationship between regions of the emotional sound network. This method correlates the time-series of different ROIs with a time lag allowing inferring causal relationship between the regions.

### *3.3.2 Applications*

The results of the studies included in this thesis are of special interest for specific populations of patients. I will focus here on two populations, namely autistic individuals and unilateral auditory spatial neglect patients.

The processing of emotional sounds and in particular HV is acquired very early in our life (Flom and Bahrick, 2007; Grossmann, 2010; Locke, 1993), with a discrimination of affective voices at around 5 months of age. However this is only the case in typically-developing children. Patients with autism spectrum disorder (ASD) show deficits in social communication and emotional voices understanding, that could implicate the voice processing network (Fan and Cheng, 2014; Gervais et al., 2004). ASD individuals exhibit enhanced processing of low-level auditory stimuli but decreased processing of complex stimuli, such as speech (Boddaert et al., 2004; O'Connor, 2012). Infant at high risk of ASD (first-degree family member diagnosed with ASD) presented an abnormal voice selectivity, with few differences in the processing of HV and NV, as it should be the case (Blasi et al., 2015). In addition to abnormal activations, ASD individuals also exhibited abnormal functional auditory connectivity patterns (Linke et al., 2018). Investigation of the processing and functional connectivity of emotional sounds in a normal population and in an ASD population would be especially interesting to draw a better picture of the impaired cognitive and neural correlates of auditory processing in ASD. This knowledge could help to develop early-stage diagnostic tools with imaging methods and cognitive individual-based therapies.

Unilateral auditory spatial neglect patients show a lack of attention to the hemispace contralesional to a brain damage (Heilman et al., 2000; Mesulam, 1999). An emotional advantage was described in this specific population of patients, using emotional faces (Domínguez-Borràs et al., 2012) or affective voices (Grandjean et al., 2008). These studies demonstrated a better rate of detection, and thus a decreased extinction rate, for emotional stimuli compared to neutral stimuli presented in the contralesional hemispace. Our finding of the enhanced processing of positive HV located on the left space could be of interest for unilateral auditory spatial neglect patients with left auditory space neglect. The use of this specific category of sounds could lead to an even decreased extinction rate. Studies described that visual intervention (prismatic adaptation; Held et al., 1966) could improve auditory extinction in specific tasks (Jacquin-Courtois et al., 2010; Tissieres et al., 2017) but it is unknown if this effect could be implemented in the other direction and thus generalize the decreased extinction rate to visual or to multi-modal neglect patients.

# 4 Bibliography

Adriani, M., Maeder, P., Meuli, R., Thiran, A.B., Frischknecht, R., Villemure, J.-G., Mayer, J., Annoni, J.-M., Bogousslavsky, J., Fornari, E., Thiran, J.-P., Clarke, S., 2003. Sound recognition and localization in man: specialized cortical networks and effects of acute circumscribed lesions. Exp. Brain Res. 153, 591–604. https://doi.org/10.1007/s00221-003-1616-0

Aeschlimann, M., Knebel, J.-F., Murray, M.M., Clarke, S., 2008. Emotional Pre-eminence of Human Vocalizations. Brain Topogr. 20, 239–248. https://doi.org/10.1007/s10548-008-0051-8

Aglieri, V., Chaminade, T., Takerkart, S., Belin, P., 2018. Functional connectivity within the voice perception network and its behavioural relevance. NeuroImage 183, 356–365. https://doi.org/10.1016/j.neuroimage.2018.08.011

Agus, T.R., Paquette, S., Suied, C., Pressnitzer, D., Belin, P., 2017. Voice selectivity in the temporal voice area despite matched low-level acoustic cues. Sci. Rep. 7, 11526. https://doi.org/10.1038/s41598-017-11684-1

Altmann, C.F., Bledowski, C., Wibral, M., Kaiser, J., 2007. Processing of location and pattern changes of natural sounds in the human auditory cortex. NeuroImage 35, 1192–1200. https://doi.org/10.1016/j.neuroimage.2007.01.007

Altmann, C.F., Nakata, H., Noguchi, Y., Inui, K., Hoshiyama, M., Kaneoke, Y., Kakigi, R., 2008. Temporal Dynamics of Adaptation to Natural Sounds in the Human Auditory Cortex. Cereb. Cortex 18, 1350–1360. https://doi.org/10.1093/cercor/bhm166

Amaral, n.d. Anatomical organization of the primate amygdaloid complex | Scinapse | Academic search engine for paper [WWW Document]. Scinapse. URL https://scinapse.io/papers/45373297 (accessed 5.3.19).

Amunts, K., Kedo, O., Kindler, M., Pieperhoff, P., Mohlberg, H., Shah, N.J., Habel, U., Schneider, F., Zilles, K., 2005. Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. Anat. Embryol. (Berl.) 210, 343–352. https://doi.org/10.1007/s00429-005-0025-5

Anderson, A.K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D.G., Glover, G., Gabrieli, J.D.E., Sobel, N., 2003. Dissociated neural representations of intensity and valence in human olfaction. Nat. Neurosci. 6, 196–202. https://doi.org/10.1038/nn1001

Aubé, W., Angulo-Perkins, A., Peretz, I., Concha, L., Armony, J.L., 2015. Fear across the senses: brain responses to music, vocalizations and facial expressions. Soc. Cogn. Affect. Neurosci. 10, 399–407. https://doi.org/10.1093/scan/nsu067

Ball, T., Rahm, B., Eickhoff, S.B., Schulze-Bonhage, A., Speck, O., Mutschler, I., 2007. Response Properties of Human Amygdala Subregions: Evidence Based on Functional MRI Combined with Probabilistic Anatomical Maps. PLOS ONE 2, e307. https://doi.org/10.1371/journal.pone.0000307

Barron, H.C., Garvert, M.M., Behrens, T.E.J., 2016. Repetition suppression: a means to index neural representations using BOLD? Philos. Trans. R. Soc. B Biol. Sci. 371. https://doi.org/10.1098/rstb.2015.0355

Baumann, S., Petkov, C.I., Griffiths, T.D., 2013. A unified framework for the organization of the primate auditory cortex. Front. Syst. Neurosci. 7. https://doi.org/10.3389/fnsys.2013.00011

Belin, P., Fecteau, S., Bédard, C., 2004. Thinking the voice: neural correlates of voice perception. Trends Cogn. Sci. 8, 129–135. https://doi.org/10.1016/j.tics.2004.01.008

Belin, P., Zatorre, R.J., Ahad, P., 2002. Human temporal-lobe response to vocal sounds. Cogn. Brain Res. 13, 17–26. https://doi.org/10.1016/S0926-6410(01)00084-2

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312. https://doi.org/10.1038/35002078

Benarroch, E.E., 2014. The amygdala Functional organization and involvement in neurologic disorders. Neurology 10.1212/WNL.0000000000001171. https://doi.org/10.1212/WNL.0000000000001171

Benner, J., Wengenroth, M., Reinhardt, J., Stippich, C., Schneider, P., Blatow, M., 2017. Prevalence and function of Heschl's gyrus morphotypes in musicians. Brain Struct. Funct. 222, 3587–3603. https://doi.org/10.1007/s00429-017-1419-x

Bergerbest, D., Ghahremani, D.G., Gabrieli, J.D.E., 2004. Neural Correlates of Auditory Repetition Priming: Reduced fMRI Activation in the Auditory Cortex. J. Cogn. Neurosci. 16, 966–977. https://doi.org/10.1162/0898929041502760

Bestelmeyer, P.E.G., Kotz, S.A., Belin, P., 2017. Effects of emotional valence and arousal on the voice perception network. Soc. Cogn. Affect. Neurosci. 12, 1351–1358. https://doi.org/10.1093/scan/nsx059

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human Temporal Lobe Activation by Speech and Nonspeech Sounds. Cereb. Cortex 10, 512–528. https://doi.org/10.1093/cercor/10.5.512

Blasi, A., Lloyd-Fox, S., Sethna, V., Brammer, M.J., Mercure, E., Murray, L., Williams, S.C.R., Simmons, A., Murphy, D.G.M., Johnson, M.H., 2015. Atypical processing of voice sounds in infants at risk for autism spectrum disorder. Cortex 71, 122–133. https://doi.org/10.1016/j.cortex.2015.06.015

Boddaert, N., Chabane, N., Belin, P., Bourgeois, M., Royer, V., Barthelemy, C., Mouren-Simeoni, M.-C., Philippe, A., Brunelle, F., Samson, Y., Zilbovicius, M., 2004. Perception of Complex Sounds in Autism: Abnormal Auditory Cortical Processing in Children. Am. J. Psychiatry 161, 2117–2120. https://doi.org/10.1176/appi.ajp.161.11.2117

Bodin, C., Takerkart, S., Belin, P., Coulon, O., 2017. Anatomo-functional correspondence in the superior temporal sulcus. Brain Struct. Funct. 1–12. https://doi.org/10.1007/s00429-017-1483-2

Bonnet, L., Comte, A., Tatu, L., Millot, J., Moulin, T., Medeiros de Bustos, E., 2015. The role of the amygdala in the perception of positive emotions: an "intensity detector." Front. Behav. Neurosci. 9, 178. https://doi.org/10.3389/fnbeh.2015.00178

Boubela, R.N., Kalcher, K., Huf, W., Seidel, E.-M., Derntl, B., Pezawas, L., Našel, C., Moser, E., 2015. fMRI measurements of amygdala activation are confounded by stimulus correlated signal fluctuation in nearby veins draining distant brain regions. Sci. Rep. 5, 10499. https://doi.org/10.1038/srep10499

Bourquin, N.M.-P., Murray, M.M., Clarke, S., 2013. Location-independent and location-linked representations of sound objects. NeuroImage 73, 40–49. https://doi.org/10.1016/j.neuroimage.2013.01.026

Bradley, M.M., Lang, P.J., 2000. Affective reactions to acoustic stimuli. Psychophysiology 37, 204–215. https://doi.org/10.1111/1469-8986.3720204

Bradley, M.M., Miccoli, L., Escrig, M.A., Lang, P.J., 2008. The pupil as a measure of emotional arousal and autonomic activation. Psychophysiology 45, 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Brefczynski-Lewis, J.A., Lewis, J.W., n.d. Auditory object perception: A neurobiological model and prospective review. Neuropsychologia. https://doi.org/10.1016/j.neuropsychologia.2017.04.034

Brouwer, A.-M., Van Wouwe, N., Muehl, C., Van Erp, J.B.F., Toet, A., 2013. Perceiving blocks of emotional pictures and sounds: effects on physiological variables. Front. Hum. Neurosci. 7, 295. https://doi.org/10.3389/fnhum.2013.00295

Brugge, J.F., 2013. Chapter 2 - Anatomy and physiology of auditory pathways and cortex, in: Celesia, G.G. (Ed.), Handbook of Clinical Neurophysiology, Disorders of Peripheral and Central Auditory Processing. Elsevier, pp. 25–59. https://doi.org/10.1016/B978-0-7020-5310-8.00002-8

Budinger, E., Heil, P., 2006. Anatomy of the auditory cortex. pp. 91–113.

Bzdok, D., Laird, A.R., Zilles, K., Fox, P.T., Eickhoff, S.B., 2013. An investigation of the structural, connectional, and functional subspecialization in the human amygdala. Hum. Brain Mapp. 34, 3247–3266. https://doi.org/10.1002/hbm.22138

Cammoun, L., Thiran, J.P., Griffa, A., Meuli, R., Hagmann, P., Clarke, S., 2015. Intrahemispheric cortico-cortical connections of the human auditory cortex. Brain Struct. Funct. 220, 3537–3553. https://doi.org/10.1007/s00429-014-0872-z

Canli, T., Zhao, Z., Desmond, J.E., Kang, E., Gross, J., Gabrieli, J.D.E., 2001. Canli T, Zhao Z, Desmond EJ, Kang E, Gross J, Gabrieli JDE. An fMRI study of personality influences on brain reactivity to emotional stimuli. ResearchGate 115, 33–42. https://doi.org/10.1037/0735-7044.115.1.33

Chevillet, M., Riesenhuber, M., Rauschecker, J.P., 2011. Functional Correlates of the Anterolateral Processing Hierarchy in Human Auditory Cortex. J. Neurosci. 31, 9345–9352. https://doi.org/10.1523/JNEUROSCI.1448-11.2011

Clarke, S., Bellmann, A., Meuli, R.A., Assal, G., Steck, A.J., 2000. Auditory agnosia and auditory spatial deficits following left hemispheric lesions: evidence for distinct processing pathways. Neuropsychologia 38, 797–807. https://doi.org/10.1016/S0028-3932(99)00141-4

Clarke, S., Geiser, E., 2015. Roaring lions and chirruping lemurs: How the brain encodes sound objects in space. Neuropsychologia 75, 304–313. https://doi.org/10.1016/j.neuropsychologia.2015.06.012

Clarke, S., Morosan, P., 2012. Architecture, connectivity and transmitter receptors of human auditory cortex, in: In Poeppel, D., Overath, T., Popper, A. N., and Fay, R.R. (Eds). (2012). Human Auditory Cortex. pp. 11–38.

Clarke, S., Thiran, A.B., Maeder, P., Adriani, M., Vernet, O., Regli, L., Cuisenaire, O., Thiran, J.-P., 2002. What and Where in human audition: selective deficits following focal hemispheric lesions. Exp. Brain Res. 147, 8–15. https://doi.org/10.1007/s00221-002-1203-9

Cohen, Y.E., Bennur, S., Christison-Lagay, K., Gifford, A.M., Tsunada, J., 2016. Functional Organization of the Ventral Auditory Pathway, in: Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing. Springer, Cham, pp. 381–388. https://doi.org/10.1007/978-3-319-25474-6_40

Costa, V.D., Lang, P.J., Sabatinelli, D., Versace, F., Bradley, M.M., 2010. Emotional Imagery: Assessing Pleasure and Arousal in the Brain's Reward Circuitry. Hum. Brain Mapp. 31, 1446–1457. https://doi.org/10.1002/hbm.20948

Da Costa, S., Clarke, S., Crottaz-Herbette, S., 2018. Keeping track of sound objects in space: The contribution of early-stage auditory areas. Hear. Res., International

Conference on Auditory Cortex 2017 366, 17–31. https://doi.org/10.1016/j.heares.2018.03.027

Da Costa, S., Bourquin, N.M.-P., Knebel, J.-F., Saenz, M., Zwaag, W. van der, Clarke, S., 2015. Representation of Sound Objects within Early-Stage Auditory Areas: A Repetition Effect Study Using 7T fMRI. PLOS ONE 10, e0124072. https://doi.org/10.1371/journal.pone.0124072

Da Costa, S., Zwaag, W. van der, Marques, J.P., Frackowiak, R.S.J., Clarke, S., Saenz, M., 2011. Human Primary Auditory Cortex Follows the Shape of Heschl's Gyrus. J. Neurosci. 31, 14067–14075. https://doi.org/10.1523/JNEUROSCI.2000-11.2011

Dalgleish, T., 2004. The emotional brain. Nat. Rev. Neurosci. 5, 583–589. https://doi.org/10.1038/nrn1432

De Martino, F., Yacoub, E., Kemper, V., Moerel, M., Uludag, K., De Weerd, P., Ugurbil, K., Goebel, R., Formisano, E., 2017. The impact of ultra-high field MRI on cognitive and computational neuroimaging. NeuroImage. https://doi.org/10.1016/j.neuroimage.2017.03.060

De Meo, R., Bourquin, N.M.-P., Knebel, J.-F., Murray, M.M., Clarke, S., 2015. From bird to sparrow: Learning-induced modulations in fine-grained semantic discrimination. NeuroImage 118, 163–173. https://doi.org/10.1016/j.neuroimage.2015.05.091

Doehrmann, O., Naumer, M.J., Volz, S., Kaiser, J., Altmann, C.F., 2008. Probing category selectivity for environmental sounds in the human auditory brain. Neuropsychologia 46, 2776–2786. https://doi.org/10.1016/j.neuropsychologia.2008.05.011

Domínguez-Borràs, J., Saj, A., Armony, J.L., Vuilleumier, P., 2012. Emotional processing and its impact on unilateral neglect and extinction. Neuropsychologia, Special Issue: Spatial Neglect and Attention 50, 1054–1071. https://doi.org/10.1016/j.neuropsychologia.2012.03.003

Duerden, E.G., Arsalidou, M., Lee, M., Taylor, M.J., 2013. Lateralization of affective processing in the insula. NeuroImage 78, 159–175. https://doi.org/10.1016/j.neuroimage.2013.04.014

Dumoulin, S.O., Fracasso, A., van der Zwaag, W., Siero, J.C.W., Petridou, N., 2017. Ultra-high field MRI: Advancing systems neuroscience towards mesoscopic human brain function. NeuroImage. https://doi.org/10.1016/j.neuroimage.2017.01.028

Engel, L.R., Frum, C., Puce, A., Walker, N.A., Lewis, J.W., 2009. Different categories of living and non-living sound-sources activate distinct cortical networks. NeuroImage 47, 1778–1791. https://doi.org/10.1016/j.neuroimage.2009.05.041

Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., Grodd, W., Wildgruber, D., 2006. Effects of prosodic emotional intensity on activation of associative auditory cortex: NeuroReport 17, 249–253. https://doi.org/10.1097/01.wnr.0000199466.32036.5d

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., Vuilleumier, P., 2012. Emotional Voice Areas: Anatomic Location, Functional Properties, and Structural Connections Revealed by Combined fMRI/DTI. Cereb. Cortex 22, 191–200. https://doi.org/10.1093/cercor/bhr113

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., Wildgruber, D., 2008. Differential Influences of Emotion, Task, and Novelty on Brain Regions Underlying the Processing of Speech Melody. J. Cogn. Neurosci. 21, 1255–1268. https://doi.org/10.1162/jocn.2009.21099

Fan, Y.-T., Cheng, Y., 2014. Atypical Mismatch Negativity in Response to Emotional Voices in People with Autism Spectrum Conditions. PLOS ONE 9, e102471. https://doi.org/10.1371/journal.pone.0102471

Fecteau, S., Belin, P., Joanette, Y., Armony, J.L., 2007. Amygdala responses to nonlinguistic emotional vocalizations. NeuroImage 36, 480–487. https://doi.org/10.1016/j.neuroimage.2007.02.043

Flom, R., Bahrick, L.E., 2007. The Development of Infant Discrimination of Affect in Multimodal and Unimodal Stimulation: The Role of Intersensory Redundancy. Dev. Psychol. 43, 238–252. https://doi.org/10.1037/0012-1649.43.1.238

Formisano, E., Kim, D.-S., Salle, F.D., Moortele, P.-F. van de, Ugurbil, K., Goebel, R., 2003. Mirror-Symmetric Tonotopic Maps in Human Primary Auditory Cortex. Neuron 40, 859–869. https://doi.org/10.1016/S0896-6273(03)00669-X

Frühholz, S., Grandjean, D., 2013. Amygdala subregions differentially respond and rapidly adapt to threatening voices. Cortex 49, 1394–1403. https://doi.org/10.1016/j.cortex.2012.08.003

Frühholz, S., Grandjean, D., 2012. Towards a fronto-temporal neural network for the decoding of angry vocal expressions. NeuroImage 62, 1658–1666. https://doi.org/10.1016/j.neuroimage.2012.06.015

Frühholz, S., Schlegel, K., Grandjean, D., 2017. Amygdala structure and core dimensions of the affective personality. Brain Struct. Funct. 1–11. https://doi.org/10.1007/s00429-017-1444-9

Frühholz, S., Trost, W., Kotz, S.A., 2016. The sound of emotions—Towards a unifying neural network perspective of affective sound processing. Neurosci. Biobehav. Rev. 68, 96–110. https://doi.org/10.1016/j.neubiorev.2016.05.002

Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., Barthélémy, C., Brunelle, F., Samson, Y., Zilbovicius, M., 2004. Abnormal cortical voice processing in autism. Nat. Neurosci. 7, 801. https://doi.org/10.1038/nn1291

Giordano, B.L., Pernet, C., Charest, I., Belizaire, G., Zatorre, R.J., Belin, P., 2014. Automatic domain-general processing of sound source identity in the left posterior middle frontal gyrus. Cortex 58, 170–185. https://doi.org/10.1016/j.cortex.2014.06.005

Gomez, P., Zimmermann, P., Guttormsen-Schär, S., Danuser, B., 2005. Respiratory responses associated with affective processing of film stimuli. Biol. Psychol. 68, 223–235. https://doi.org/10.1016/j.biopsycho.2004.06.003

Grandjean, D., Sander, D., Lucas, N., Scherer, K.R., Vuilleumier, P., 2008. Effects of emotional prosody on auditory extinction for voices in patients with spatial neglect. Neuropsychologia 46, 487–496. https://doi.org/10.1016/j.neuropsychologia.2007.08.025

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., Vuilleumier, P., 2005. The voices of wrath: brain responses to angry prosody in meaningless speech. Nat. Neurosci. 8, 145–146. https://doi.org/10.1038/nn1392

Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn. Sci. 10, 14–23. https://doi.org/10.1016/j.tics.2005.11.006

Grisendi T, Reynaud O, Clarke S, Da Costa S (2019) Processing pathways for emotional vocalizations. Brain Struct Funct.; In revision.

Grossmann, T., 2010. The development of emotion perception in face and voice during infancy. Restor. Neurol. Neurosci. 28, 219–236. https://doi.org/10.3233/RNN-2010-0499

Grothe, B., Pecka, M., McAlpine, D., 2010. Mechanisms of Sound Localization in Mammals. Physiol. Rev. 90, 983–1012. https://doi.org/10.1152/physrev.00026.2009

H Kaas, J., A Hackett, T., 2005. Subdivisions and connections of the auditory cortex in primates: A working model. pp. 7–26.

Hackett, T.A., Kaas, J.H., 2004. Auditory Cortex in Primates: Functional Subdivisions and Processing Streams, in: The Cognitive Neurosciences, 3rd Ed. MIT Press, Cambridge, MA, US, pp. 215–232.

Hall, D.A., Johnsrude, I.S., Haggard, M.P., Palmer, A.R., Akeroyd, M.A., Summerfield, A.Q., 2002. Spectral and Temporal Processing in Human Auditory Cortex. Cereb. Cortex 12, 140–149. https://doi.org/10.1093/cercor/12.2.140

Heilman, K.M., Valenstein, E., Watson, R.T., 2000. Neglect and Related Disorders. Semin. Neurol. 20, 463–470. https://doi.org/10.1055/s-2000-13179

Held, R., Efstathiou, A., Greene, M., 1966. Adaptation to displaced and delayed visual feedback from the hand. J. Exp. Psychol. 72, 887–891. https://doi.org/10.1037/h0023868

Humphries, C., Liebenthal, E., Binder, J.R., 2010. Tonotopic organization of human auditory cortex. NeuroImage 50, 1202–1211. https://doi.org/10.1016/j.neuroimage.2010.01.046

Hurlemann, R., Rehme, A.K., Diessel, M., Kukolja, J., Maier, W., Walter, H., Cohen, M.X., 2008. Segregating intra-amygdalar responses to dynamic facial emotion with cytoarchitectonic maximum probability maps. J. Neurosci. Methods 172, 13–20. https://doi.org/10.1016/j.jneumeth.2008.04.004

Jacquin-Courtois, S., Rode, G., Pavani, F., O’Shea, J., Giard, M.H., Boisson, D., Rossetti, Y., 2010. Effect of prism adaptation on left dichotic listening deficit in neglect patients: glasses to hear better? Brain 133, 895–908. https://doi.org/10.1093/brain/awp327

Johnstone, T., Reekum, C.M. van, Oakes, T.R., Davidson, R.J., 2006. The voice of emotion: an FMRI study of neural responses to angry and happy vocal expressions. Soc. Cogn. Affect. Neurosci. 1, 242–249. https://doi.org/10.1093/scan/nsl027

Kasper, L., Bollmann, S., Diaconescu, A.O., Hutton, C., Heinzle, J., Iglesias, S., Hauser, T.U., Sebold, M., Manjaly, Z.-M., Pruessmann, K.P., Stephan, K.E., 2017. The PhysIO Toolbox for Modeling Physiological Noise in fMRI Data. J. Neurosci. Methods 276, 56–72. https://doi.org/10.1016/j.jneumeth.2016.10.019

Kim, M.J., Loucks, R.A., Palmer, A.L., Brown, A.C., Solomon, K.M., Marchante, A.N., Whalen, P.J., 2011. The structural and functional connectivity of the amygdala: From normal emotion to pathological anxiety. Behav. Brain Res. 223, 403–410. https://doi.org/10.1016/j.bbr.2011.04.025

Koelsch, S., 2010. Towards a neural basis of music-evoked emotions. Trends Cogn. Sci. 14, 131–137. https://doi.org/10.1016/j.tics.2010.01.002

Kryklywy, J.H., Macpherson, E.A., Greening, S.G., Mitchell, D.G.V., 2013. Emotion modulates activity in the ‘what’ but not ‘where’ auditory processing pathway. NeuroImage 82, 295–305. https://doi.org/10.1016/j.neuroimage.2013.05.051

Kumar, S., Kriegstein, K. von, Friston, K., Griffiths, T.D., 2012. Features versus Feelings: Dissociable Representations of the Acoustic Features and Valence of Aversive

Sounds. J. Neurosci. 32, 14184–14192. https://doi.org/10.1523/JNEUROSCI.1759-12.2012

Kumar, S., Stephan, K.E., Warren, J.D., Friston, K.J., Griffiths, T.D., 2007. Hierarchical Processing of Auditory Objects in Humans. PLOS Comput. Biol. 3, e100. https://doi.org/10.1371/journal.pcbi.0030100

Labar, K.S., Gitelman, D.R., Mesulam, M.M., Parrish, T.B., 2001. Impact of signal-to-noise on functional Mri of the human amygdala. Neuroreport 12, 3461–3464.

Laeger, I., Dobel, C., Dannlowski, U., Kugel, H., Grotegerd, D., Kissler, J., Keuper, K., Eden, A., Zwitserlood, P., Zwanzger, P., 2012. Amygdala responsiveness to emotional words is modulated by subclinical anxiety and depression. Behav. Brain Res. 233, 508–516. https://doi.org/10.1016/j.bbr.2012.05.036

Latinus, M., McAleer, P., Bestelmeyer, P.E.G., Belin, P., 2013. Norm-Based Coding of Voice Identity in Human Auditory Cortex. Curr. Biol. 23, 1075–1080. https://doi.org/10.1016/j.cub.2013.04.055

Leaver, A.M., Rauschecker, J.P., 2016. Functional Topography of Human Auditory Cortex. J. Neurosci. 36, 1416–1428. https://doi.org/10.1523/JNEUROSCI.0226-15.2016

Leaver, A.M., Rauschecker, J.P., 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J. Neurosci. Off. J. Soc. Neurosci. 30, 7604–7612. https://doi.org/10.1523/JNEUROSCI.0296-10.2010

LeDoux, J., 2007. The amygdala. Curr. Biol. 17, R868–R874. https://doi.org/10.1016/j.cub.2007.08.005

LeDoux, J.E., 2000. Emotion Circuits in the Brain. Annu. Rev. Neurosci. 23, 155–184. https://doi.org/10.1146/annurev.neuro.23.1.155

LeDoux, J.E., Farb, C., Ruggiero, D.A., 1990. Topographic organization of neurons in the acoustic thalamus that project to the amygdala. J. Neurosci. 10, 1043–1054. https://doi.org/10.1523/JNEUROSCI.10-04-01043.1990

Leech, R., Saygin, A.P., 2011. Distributed processing and cortical specialization for speech and environmental sounds in human temporal cortex. Brain Lang. 116, 83–90. https://doi.org/10.1016/j.bandl.2010.11.001

Leitman, D.I., Wolf, D.H., Ragland, J.D., Laukka, P., Loughead, J., Valdez, J.N., Javitt, D.C., Turetsky, B.I., Gur, R.C., 2010. "It's Not What You Say, But How You Say it": A Reciprocal Temporo-frontal Network for Affective Prosody. Front. Hum. Neurosci. 4. https://doi.org/10.3389/fnhum.2010.00019

Lewis, J.W., Brefczynski, J.A., Phinney, R.E., Janik, J.J., DeYoe, E.A., 2005. Distinct Cortical Pathways for Processing Tool versus Animal Sounds. J. Neurosci. 25, 5148–5158. https://doi.org/10.1523/JNEUROSCI.0419-05.2005

Lewis, J.W., Wightman, F.L., Brefczynski, J.A., Phinney, R.E., Binder, J.R., DeYoe, E.A., 2004. Human Brain Regions Involved in Recognizing Environmental Sounds. Cereb. Cortex 14, 1008–1021. https://doi.org/10.1093/cercor/bhh061

Liebenthal, E., Silbersweig, D.A., Stern, E., 2016. The Language, Tone and Prosody of Emotions: Neural Substrates and Dynamics of Spoken-Word Emotion Perception. Front. Neurosci. 10. https://doi.org/10.3389/fnins.2016.00506

Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E., Barrett, L.F., 2012. The brain basis of emotion: A meta-analytic review. Behav. Brain Sci. 35, 121–143. https://doi.org/10.1017/S0140525X11000446

Linke, A.C., Jao Keehn, R.J., Pueschel, E.B., Fishman, I., Müller, R.-A., 2018. Children with ASD show links between aberrant sound processing, social symptoms, and atypical auditory interhemispheric and thalamocortical functional connectivity. Dev.

Cogn. Neurosci., Autism Spectrum Condition – understanding sensory and social features 29, 117–126. https://doi.org/10.1016/j.dcn.2017.01.007

Locke, J.L., 1993. The child's path to spoken language, The child's path to spoken language. Harvard University Press, Cambridge, MA, US.

Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A., 2001. Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150. https://doi.org/10.1038/35084005

Lohmann, G., Margulies, D.S., Horstmann, A., Pleger, B., Lepsien, J., Goldhahn, D., Schloegl, H., Stumvoll, M., Villringer, A., Turner, R., 2010. Eigenvector Centrality Mapping for Analyzing Connectivity Patterns in fMRI Data of the Human Brain. PLOS ONE 5, e10232. https://doi.org/10.1371/journal.pone.0010232

Lucia, M.D., Cocchi, L., Martuzzi, R., Meuli, R.A., Clarke, S., Murray, M.M., 2010. Perceptual and Semantic Contributions to Repetition Priming of Environmental Sounds. Cereb. Cortex 20, 1676–1684. https://doi.org/10.1093/cercor/bhp230

Maeder, P.P., Meuli, R.A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J.-P., Pittet, A., Clarke, S., 2001. Distinct Pathways Involved in Sound Recognition and Localization: A Human fMRI Study. NeuroImage 14, 802–816. https://doi.org/10.1006/nimg.2001.0888

Martin-Soelch, C., Stöcklin, M., Dammann, G., Opwis, K., Seifritz, E., 2006. Anxiety trait modulates psychophysiological reactions, but not habituation processes related to affective auditory stimuli. Int. J. Psychophysiol. 61, 87–97. https://doi.org/10.1016/j.ijpsycho.2005.07.009

Mathiak, K.A., Zvyagintsev, M., Ackermann, H., Mathiak, K., 2012. Lateralization of amygdala activation in fMRI may depend on phase-encoding polarity. Magn. Reson. Mater. Phys. Biol. Med. 25, 177–182. https://doi.org/10.1007/s10334-011-0285-4

Mesulam, M.M., 1999. Spatial attention and neglect: parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. Philos. Trans. R. Soc. B Biol. Sci. 354, 1325–1346.

Moerel, M., De Martino, F., Formisano, E., 2014. An anatomical and functional topography of human auditory cortical areas. Front. Neurosci. 8. https://doi.org/10.3389/fnins.2014.00225

Moerel, M., Martino, F.D., Formisano, E., 2012. Processing of Natural Sounds in Human Auditory Cortex: Tonotopy, Spectral Tuning, and Relation to Voice Sensitivity. J. Neurosci. 32, 14205–14216. https://doi.org/10.1523/JNEUROSCI.1388-12.2012

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., Zilles, K., 2001. Human Primary Auditory Cortex: Cytoarchitectonic Subdivisions and Mapping into a Spatial Reference System. NeuroImage 13, 684–701. https://doi.org/10.1006/nimg.2000.0715

Morris, J.S., Friston, K.J., Büchel, C., Frith, C.D., Young, A.W., Calder, A.J., Dolan, R.J., 1998. A neuromodulatory role for the human amygdala in processing emotional facial expressions. Brain 121, 47–57. https://doi.org/10.1093/brain/121.1.47

Morris, J.S., Scott, S.K., Dolan, R.J., 1999. Saying it with feeling: neural responses to emotional vocalizations. Neuropsychologia 37, 1155–1163. https://doi.org/10.1016/S0028-3932(99)00015-9

Murray, M.M., Camen, C., Andino, S.L.G., Bovet, P., Clarke, S., 2006. Rapid Brain Discrimination of Sounds of Objects. J. Neurosci. 26, 1293–1302. https://doi.org/10.1523/JNEUROSCI.4511-05.2006

Murray, M.M., Camen, C., Spierer, L., Clarke, S., 2008. Plasticity in representations of environmental sounds revealed by electrical neuroimaging. NeuroImage 39, 847–856. https://doi.org/10.1016/j.neuroimage.2007.09.002

Narsude, M., Gallichan, D., van der Zwaag, W., Gruetter, R., Marques, J.P., 2016. Three-dimensional echo planar imaging with controlled aliasing: A sequence for high temporal resolution functional MRI. Magn. Reson. Med. 75, 2350–2361. https://doi.org/10.1002/mrm.25835

O'Connor, K., 2012. Auditory processing in autism spectrum disorder: A review. Neurosci. Biobehav. Rev. 36, 836–854. https://doi.org/10.1016/j.neubiorev.2011.11.008

O'Doherty, J., Rolls, E.T., Francis, S., Bowtell, R., McGlone, F., 2001. Representation of Pleasant and Aversive Taste in the Human Brain. J. Neurophysiol. 85, 1315–1321.

Pannese, A., Grandjean, D., Frühholz, S., 2015. Subcortical processing in auditory communication. Hear. Res. 328, 67–77. https://doi.org/10.1016/j.heares.2015.07.003

Partala, T., Surakka, V., 2003. Pupil size variation as an indication of affective processing. Int. J. Hum.-Comput. Stud., Applications of Affective Computing in Human-Computer Interaction 59, 185–198. https://doi.org/10.1016/S1071-5819(03)00017-X

Pedersen, W.S., Balderston, N.L., Miskovich, T.A., Belleau, E.L., Helmstetter, F.J., Larson, C.L., 2017. The effects of stimulus novelty and negativity on BOLD activity in the amygdala, hippocampus, and bed nucleus of the stria terminalis. Soc. Cogn. Affect. Neurosci. 12, 748–757. https://doi.org/10.1093/scan/nsw178

Peelen, M.V., Atkinson, A.P., Vuilleumier, P., 2010. Supramodal Representations of Perceived Emotions in the Human Brain. J. Neurosci. 30, 10127–10134. https://doi.org/10.1523/JNEUROSCI.2161-10.2010

Pell, M.D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., Rigoulot, S., 2015. Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. Biol. Psychol. 111, 14–25. https://doi.org/10.1016/j.biopsycho.2015.08.008

Pernet, C.R., McAleer, P., Latinus, M., Gorgolewski, K.J., Charest, I., Bestelmeyer, P.E.G., Watson, R.H., Fleming, D., Crabbe, F., Valdes-Sosa, M., Belin, P., 2015. The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. NeuroImage 119, 164–174. https://doi.org/10.1016/j.neuroimage.2015.06.050

Phan, K.L., Wager, T., Taylor, S.F., Liberzon, I., 2002. Functional Neuroanatomy of Emotion: A Meta-Analysis of Emotion Activation Studies in PET and fMRI. NeuroImage 16, 331–348. https://doi.org/10.1006/nimg.2002.1087

Phillips, M.L., Young, A.W., Scott, S.K., Calder, A.J., Andrew, C., Giampietro, V., Williams, S.C.R., Bullmore, E.T., Brammer, M., Gray, J.A., 1998. Neural responses to facial and vocal expressions of fear and disgust. Proc. R. Soc. Lond. B Biol. Sci. 265, 1809–1817. https://doi.org/10.1098/rspb.1998.0506

Plichta, M.M., Gerdes, A.B.M., Alpers, G.W., Harnisch, W., Brill, S., Wieser, M.J., Fallgatter, A.J., 2011. Auditory cortex activation is modulated by emotion: A functional near-infrared spectroscopy (fNIRS) study. NeuroImage 55, 1200–1207. https://doi.org/10.1016/j.neuroimage.2011.01.011

Poldrack, R.A., Mumford, J.A., Nichols, T.E., 2011. Handbook of Functional MRI Data Analysis. Cambridge University Press.

Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 12, 718–724. https://doi.org/10.1038/nn.2331

Rauschecker, J.P., Tian, B., 2000. Mechanisms and streams for processing of “what” and “where” in auditory cortex. Proc. Natl. Acad. Sci. 97, 11800–11806. https://doi.org/10.1073/pnas.97.22.11800

Reynaud, O., Jorge, J., Gruetter, R., Marques, J.P., Zwaag, W. van der, 2017. Influence of physiological noise on accelerated 2D and 3D resting state functional MRI data at 7 T. Magn. Reson. Med. 78, 888–896. https://doi.org/10.1002/mrm.26823

Rivier, F., Clarke, S., 1997. Cytochrome Oxidase, Acetylcholinesterase, and NADPH-Diaphorase Staining in Human Supratemporal and Insular Cortex: Evidence for Multiple Auditory Areas. NeuroImage 6, 288–304. https://doi.org/10.1006/nimg.1997.0304

Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. NeuroImage 25, 230–242. https://doi.org/10.1016/j.neuroimage.2004.11.017

Saenz, M., Langers, D.R.M., 2014. Tonotopic mapping of human auditory cortex. Hear. Res., Human Auditory NeuroImaging 307, 42–52. https://doi.org/10.1016/j.heares.2013.07.016

Sah, P., Faber, E.S.L., Armentia, M.L.D., Power, J., 2003. The Amygdaloid Complex: Anatomy and Physiology. Physiol. Rev. 83, 803–834. https://doi.org/10.1152/physrev.00002.2003

Sander, D., Grafman, J., Zalla, T., 2003. The human amygdala: an evolved system for relevance detection. Rev. Neurosci. 14, 303–316.

Sander, K., Scheich, H., 2005. Left Auditory Cortex and Amygdala, but Right Insula Dominance for Human Laughing and Crying. J. Cogn. Neurosci. 17, 1519–1531. https://doi.org/10.1162/089892905774597227

Sander, K., Scheich, H., 2001. Auditory perception of laughing and crying activates human amygdala regardless of attentional state. Cogn. Brain Res. 12, 181–198. https://doi.org/10.1016/S0926-6410(01)00045-3

Sauter, D.A., Eimer, M., 2009. Rapid Detection of Emotion from Human Vocalizations. J. Cogn. Neurosci. 22, 474–481. https://doi.org/10.1162/jocn.2009.21215

Schirmer, A., Gunter, T.C., 2017. Temporal signatures of processing voiceness and emotion in sound. Soc. Cogn. Affect. Neurosci. 12, 902–909. https://doi.org/10.1093/scan/nsx020

Schirmer, A., Kotz, S.A., 2006. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. Trends Cogn. Sci. 10, 24–30. https://doi.org/10.1016/j.tics.2005.11.009

Schneider, P., Scherg, M., Dosch, H.G., Specht, H.J., Gutschalk, A., Rupp, A., 2002. Morphology of Heschl’s gyrus reflects enhanced activation in the auditory cortex of musicians. Nat. Neurosci. 5, 688–694. https://doi.org/10.1038/nn871

Scott, S.K., 2005. Auditory processing — speech, space and auditory objects. Curr. Opin. Neurobiol., Cognitive neuroscience 15, 197–201. https://doi.org/10.1016/j.conb.2005.03.009

Sergerie, K., Chochol, C., Armony, J.L., 2008. The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies. Neurosci. Biobehav. Rev. 32, 811–830. https://doi.org/10.1016/j.neubiorev.2007.12.002

Sladky, R., Baldinger, P., Kranz, G.S., Tröstl, J., Höflich, A., Lanzenberger, R., Moser, E., Windischberger, C., 2013. High-resolution functional MRI of the human amygdala at 7 T. Eur. J. Radiol. 82, 728–733. https://doi.org/10.1016/j.ejrad.2011.09.025

Stein, M.B., Simmons, A.N., Feinstein, J.S., Paulus, M.P., 2007. Increased Amygdala and Insula Activation During Emotion Processing in Anxiety-Prone Subjects. Am. J. Psychiatry 164, 318–327. https://doi.org/10.1176/ajp.2007.164.2.318

Striem-Amit, E., Hertz, U., Amedi, A., 2011. Extensive Cochleotopic Mapping of Human Auditory Cortical Fields Obtained with Phase-Encoding fMRI. PLOS ONE 6, e17832. https://doi.org/10.1371/journal.pone.0017832

Talavage, T.M., Gonzalez-Castillo, J., Scott, S.K., 2014. Auditory neuroimaging with fMRI and PET. Hear. Res., Human Auditory NeuroImaging 307, 4–15. https://doi.org/10.1016/j.heares.2013.09.009

Talavage, T.M., Sereno, M.I., Melcher, J.R., Ledden, P.J., Rosen, B.R., Dale, A.M., 2004. Tonotopic Organization in Human Auditory Cortex Revealed by Progressions of Frequency Sensitivity. J. Neurophysiol. 91, 1282–1296. https://doi.org/10.1152/jn.01125.2002

Tardif, E., Murray, M.M., Meylan, R., Spierer, L., Clarke, S., 2006. The spatio-temporal brain dynamics of processing and integrating sound localization cues in humans. Brain Res. 1092, 161–176. https://doi.org/10.1016/j.brainres.2006.03.095

Tissieres, I., Elamly, M., Clarke, S., Crottaz-Herbette, S., 2017. For Better or Worse: The Effect of Prismatic Adaptation on Auditory Neglect [WWW Document]. Neural Plast. https://doi.org/10.1155/2017/8721240

Triantafyllou, C., Hoge, R.D., Krueger, G., Wiggins, C.J., Potthast, A., Wiggins, G.C., Wald, L.L., 2005. Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. NeuroImage 26, 243–250. https://doi.org/10.1016/j.neuroimage.2005.01.007

Tschacher, W., Schildt, M., Sander, K., 2010. Brain connectivity In listening to affective stimuli: A functional magnetic resonance imaging (fMRI) study and implications for psychotherapy. Psychother. Res. 20, 576–588. https://doi.org/10.1080/10503307.2010.493538

van der Zwaag, W., Reynaud, O., Narsude, M., Gallichan, D., Marques, J.P., 2018. High spatio-temporal resolution in functional MRI with 3D echo planar imaging using cylindrical excitation and a CAIPIRINHA undersampling pattern. Magn. Reson. Med. 79, 2589–2596. https://doi.org/10.1002/mrm.26906

van der Zwaag, W., Da Costa, S.E., Zürcher, N.R., Adams, R.B., Hadjikhani, N., 2012. A 7 Tesla fMRI Study of Amygdala Responses to Fearful Faces. Brain Topogr. 25, 125–128. https://doi.org/10.1007/s10548-012-0219-0

van der Zwaag, W., Francis, S., Head, K., Peters, A., Gowland, P., Morris, P., Bowtell, R., 2009. fMRI at 1.5, 3 and 7 T: Characterising BOLD signal changes. NeuroImage 47, 1425–1434. https://doi.org/10.1016/j.neuroimage.2009.05.015

van der Zwaag, W., Gentile, G., Gruetter, R., Spierer, L., Clarke, S., 2011. Where sound position influences sound object representations: A 7-T fMRI study. NeuroImage 54, 1803–1811. https://doi.org/10.1016/j.neuroimage.2010.10.032

Viceic, D., Fornari, E., Thiran, J.-P., Maeder, P.P., Meuli, R., Adriani, M., Clarke, S., 2006. Human auditory belt areas specialized in sound recognition: a functional magnetic resonance imaging study. Neuroreport 17, 1659–1662. https://doi.org/10.1097/01.wnr.0000239962.75943.dd

Viinikainen, M., Kätsyri, J., Sams, M., 2012. Representation of perceived sound valence in the human brain. Hum. Brain Mapp. 33, 2295–2305. https://doi.org/10.1002/hbm.21362

Vrticka, P., Sander, D., Vuilleumier, P., 2013. Lateralized interactive social content and valence processing within the human amygdala. Front. Hum. Neurosci. 6. https://doi.org/10.3389/fnhum.2012.00358

Wallace, M.N., Johnston, P.W., Palmer, A.R., 2002. Histochemical identification of cortical areas in the auditory region of the human brain. Exp. Brain Res. 143, 499–508. https://doi.org/10.1007/s00221-002-1014-z

Warren, J.D., Jennings, A.R., Griffiths, T.D., 2005. Analysis of the spectral envelope of sounds by the human brain. NeuroImage 24, 1052–1057. https://doi.org/10.1016/j.neuroimage.2004.10.031

Warren, J.D., Zielinski, B.A., Green, G.G.R., Rauschecker, J.P., Griffiths, T.D., 2002. Perception of Sound-Source Motion by the Human Brain. Neuron 34, 139–148. https://doi.org/10.1016/S0896-6273(02)00637-2

Wessinger, C.M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., Rauschecker, J.P., 2001. Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. J. Cogn. Neurosci. 13, 1–7.

Wiethoff, S. a b, Wildgruber, D. a, Grodd, W. b, Ethofer, T. a b, 2009. Response and habituation of the amygdala during processing of emotional prosody. [Miscellaneous Article]. Neuroreport 20, 1356–1360. https://doi.org/10.1097/WNR.0b013e328330eb83

Winston, J.S., Gottfried, J.A., Kilner, J.M., Dolan, R.J., 2005. Integrated Neural Representations of Odor Intensity and Affective Valence in Human Amygdala. J. Neurosci. 25, 8903–8907. https://doi.org/10.1523/JNEUROSCI.1569-05.2005

Woodson, W., Farb, C.R., Ledoux, J.E., 2000. Afferents from the auditory thalamus synapse on inhibitory interneurons in the lateral nucleus of the amygdala. Synap. N. Y. N 38, 124–137. https://doi.org/10.1002/1098-2396(200011)38:2<124::AID-SYN3>3.0.CO;2-N

Wright, C., Fischer, H., Whalen, P., McInerney, S., Shin, L., Rauch, S., 2001. Differential prefrontal cortex and amygdala habituation to repeatedly presented emotional stimuli. Neuroreport 12, 379–383.

Wright, C.I., Martis, B., Schwartz, C.E., Shin, L.M., Fischer, H. åkan, McMullin, K., Rauch, S.L., 2003. Novelty responses and differential effects of order in the amygdala, substantia innominata, and inferior temporal cortex. NeuroImage 18, 660–669. https://doi.org/10.1016/S1053-8119(02)00037-X

Zald, D.H., 2003. The human amygdala and the emotional evaluation of sensory stimuli. Brain Res. Rev. 41, 88–123. https://doi.org/10.1016/S0165-0173(02)00248-5

Zald, D.H., Pardo, J.V., 2002. The Neural Correlates of Aversive Auditory Stimulation. NeuroImage 16, 746–753. https://doi.org/10.1006/nimg.2002.1115

# 5 Research papers

## STUDY A: Processing pathways for emotional vocalizations

Tiffany Grisendi, Olivier Reynaud, Stephanie Clarke and Sandra Da Costa

Accepted in **Brain Structure and Function**

# Processing pathways for emotional vocalizations

**Tiffany Grisendi[1], Olivier Reynaud[2], Stephanie Clarke[1] and Sandra Da Costa[2]**

[1] Service de Neuropsychologie et de Neuroréhabilitation, Centre Hospitalier Universitaire Vaudois (CHUV) and University of Lausanne, 1011 Lausanne, Switzerland

[2] Centre d'Imagerie Biomédicale, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

**Corresponding author:**
Prof. Stephanie Clarke
Service de neuropsychologie et de neuroréhabilitation
CHUV | Centre hospitalier universitaire vaudois
1011 Lausanne,
Switzerland
Tel. +41 21 314 13 09
ORCID 0000-0001-9393-5718
E-mail Stephanie.Clarke@chuv.ch

Number of pages: 24
Number of figures: 4
Number of tables: 4
Number of words in Abstract: 161

**Sharing data:** The dataset of this manuscript is available upon request to the corresponding author.

## Abstract

Emotional sounds are processed within a large cortico-subcortical network, of which the auditory cortex, the voice area and the amygdala are the core regions. Using 7T fMRI we have compared the effect of emotional valence (positive, neutral, negative) and the effect of the type of environmental sounds (human vocalizations, non-vocalizations) on neural activity within individual early-stage auditory areas, the voice area and the amygdala. A 2-way ANOVA was applied to the BOLD time course within each ROI. In several early-stage auditory areas it yielded a significant main effect of vocalizations and of valence, but not a significant interaction. Significant interaction as well as significant main effects of vocalization and of valence were present in the voice area; the former was driven by a significant emotional modulation of vocalizations but not of other sounds. Within the amygdala only the main effect of valence was significant. Post hoc correlation analysis highlighted coupling between the voice area and early-stage auditory areas during the presentation of any vocalizations, and between the voice area and the right amygdala during positive vocalizations. Thus, the voice area is selectively devoted to the encoding of the emotional valence of vocalizations; it shares with several early-stage auditory areas encoding characteristics for vocalizations and with the amygdala for the emotional modulation of vocalizations. These results are indicative of a dual pathway, whereby the emotional modulation of vocalizations within the voice area integrates the input from the lateral early-stage auditory areas and from the amygdala.

**Keywords:** Human vocalizations, emotions, auditory belt areas, amygdala, voice area, 7T fMRI

## Abbreviations

| | |
|---|---|
| AI | primary auditory area |
| AMY | amygdala |
| HVN | human vocalizations with negative emotional valence |
| HVP | human vocalizations with positive emotional valence |
| HV0 | human vocalizations with neutral emotional valence |
| NVN | non-vocalizations with negative emotional valence |
| NVP | non-vocalizations with positive emotional valence |
| NV0 | non-vocalizations with neutral emotional valence |
| R | rostral (primary) auditory area |
| VA | voice area |

## Introduction

The universal nature of human emotional vocalizations and the prominent role they play in shaping social interactions (Sauter et al., 2010) warrants the highly specialized processing, which has been described in a series of seminal studies. The emotional valence of human vocalizations is processed by complex cortico-subcortical networks, which include the primary auditory cortex, the surrounding early-stage and higher-order auditory areas, the inferior frontal gyrus and the amygdala, as well as the medial frontal cortex, the insula, basal ganglia and the cerebellum (Schirmer and Kotz, 2006; Frühholz et al., 2016). As highlighted in a recent review (Frühholz et al., 2016) the amygdala, the auditory areas, including the voice area, and the inferior frontal cortex have strong functional connections and encode not only the emotional valence of non-verbal vocalizations but also of other sound categories (e. g. speech prosody: Wildgruber et al., 2009 ; music: Koelsch, 2010).

The *amygdala* (AMY) has been repeatedly shown to be involved in the processing of emotional stimuli of different sensory modalities, as demonstrated by fMRI and PET studies (Baas et al., 2004; Costafreda et al., 2008; Ball et al., 2009) and by intracranial recordings (for review Murray et al., 2014). In the auditory modality emotional modulation was investigated for human vocalizations (Morris et al., 1999; Sander and Scheich, 2005; Ethofer et al., 2006a, 2006b; Viinikainen et al., 2012; Pannese et al., 2016), including emotional prosody (reviews: Wildgruber et al., 2006; Liebenthal et al., 2016); a mixture of human vocalizations and environmental sounds (Viinikainen et al., 2012); or instrumental music (reviews: Koelsch, 2010; Frühholz et al., 2014). Cytoarchitectonically AMY is subdivided into three major nuclei groups, the lateralbasal, centromedial and superficial (Amunts et al., 2005). As demonstrated in non-human primate and non-primate species, it receives auditory input via monosynaptic afferents from the medial geniculate nucleus (Ottersen and Ben-Ari, 1979; Russchen, 1982; LeDoux et al., 1985; Shinonaga et al., 1994) and by reciprocal interconnections with parts of the auditory cortex. The latter involve in non-human primates predominantly the non-primary auditory areas and cortical regions on the postero-superior part of the temporal convexity (Price and Amaral, 1981; Yukie, 2002); in some species part of the primary auditory cortex was found to be involved as well (Reser et al., 2009). Among other wide-spread connections, AMY has also reciprocal connections with the prefrontal cortex (Ghashghaei and Barbas, 2002; Barbas, 2007) and with the mediodorsal nucleus of the thalamus (Russchen et al., 1987). Diffusion tensor imaging in humans revealed fibre tracts between AMY and medial geniculate nucleus related fibre tracts that are compatible with the connectivity described in non-human species (Keifer et al., 2015; Kamali et al., 2016). As shown in rodents, distinct amygdala circuits processes fearful and rewarding stimuli and modulate, via specific outputs, autonomic reactions (Janak and Tye, 2015).

The *voice area* (VA) was defined in the pioneering study of Belin and colleagues (Belin et al., 2000) by its stronger responses to human than animal vocalizations. It is located in the middle part of the superior temporal gyrus and sulcus. A follow-up study using the same stimulus set showed in 218 subjects that the voice-sensitive region is mostly bilateral (94% of subjects). It confirmed the location within the posterior part of the superior temporal sulcus and on the adjacent part of the convexity of the superior temporal gyrus; the rostral extension of VA was shown to reach up to the (lower) lip of the sylvian fissure and to encroach on the lateral brim of the supratemporal plane. The authors reported great inter-individual variability as to its precise anatomical location (Pernet et al., 2015). The part of the superior temporal gyrus and sulcus, where VA is located, receives auditory input from the primary auditory cortex via a cascade of cortico-cortical connections (Cammoun et al., 2015). In this respect it is similar to other auditory processing pathways in human (e. g. Kim and Knösche, 2016) and bears strong similarity to the homologous region in non-human primates (e. g. a recent comprehensive study: Scott et al., 2015). VA activity was reported to be modulated by the emotional value of vocalizations (Belin et al., 2002; Grandjean et al., 2005; Ethofer et al., 2006b, 2008, 2009b; Beaucousin et al., 2007a; Obleser et al., 2007, 2008; Bestelmeyer et al., 2017). The *emotional voice area* has been identified by its stronger response to emotional than neutral pseudo-sentences; it overlaps partially with VA and extends beyond it to the middle part of the superior temporal gyrus, Heschl's gyrus, and the antero-lateral part of the planum temporale (Ethofer et al., 2012). Thus, in addition to VA, the auditory cortex on the supratemporal plane participates significantly in the encoding of emotional vocalizations (Wildgruber et al., 2004a; Meyer et al., 2005; Dietrich et al., 2007, 2008; Leitman et al., 2010; Szameitat et al., 2010; Ethofer et al., 2012). Although multiple subregions of the superior temporal cortex are involved (Frühholz and Grandjean, 2013), none of previous studies analysed individual auditory areas, most likely because of limitations imposed by low spatial resolution. From the above studies eight used 1.5T (Belin et al., 2000, 2002; Wildgruber et al., 2004b; Grandjean et al., 2005; Ethofer et al., 2006b, 2008; Beaucousin et al., 2007b; Szameitat et al., 2010) and nine 3T fMRI (Dietrich et al., 2007; Ethofer et al., 2008, 2009a, 2012; Obleser et al., 2008; Leitman et al., 2010; Arnal et al., 2015; Bestelmeyer et al., 2017; Lavan et al., 2017), often with a voxel size of 3x3x3 mm or more (11 studies) and smoothing of more than 8 mm (12 studies). In view of the anatomical evidence described below, this type of spatial resolution does not allow to analyse individual early-stage auditory areas.

The supratemporal plane comprises several *early-stage auditory areas*, as demonstrated in histological studies (Clarke and Morosan, 2012). Investigating them with fMRI represents a challenge on three accounts. First, early-stage auditory areas tend to be relatively small (40-310 mm2; (Clarke and Morosan, 2012) and their investigation requires high spatial resolution. Second, the realignment of the supratemporal plane tends to be imprecise and landmarks, such as Heschl's gyrus, has been shown to be shifted by as much as 4 mm between individual brains of a group study (Viceic et al., 2009). Thus, a whole brain contrast aligns in a group study regions, which do not correspond across subjects to the same area. To circumvent this problem, several studies used a functional marker for the primary auditory cortex and based its identification in each individual subject on tonotopic mapping (with ultrahigh field: Formisano et al., 2003; Da Costa et al., 2011; Moerel et al., 2014), or by approximating it with Heschl's gyrus (Zilles et al., 1988; Rademacher et al., 2001; Viceic et al., 2006; van der Zwaag et al., 2011). Third, no reliable functional marker is available for individual non-primary auditory areas; they tend to be broadly tuned to complex features such as frequency, pitch, amplitude modulation or envelop (Hall et al., 2002; Rauschecker and Scott, 2009; Chevillet et al., 2011). Their characterization relies on histological criteria and was carried out in *post-mortem* material (Rivier and Clarke, 1997; Clarke and Rivier, 1998; Hackett et al., 2001; Wallace et al., 2002; Chiry et al., 2003). Their identification in activation studies can be based on Talairach coordinates published in histological studies (Viceic et al., 2006; van der Zwaag et al., 2011), preferably in combination with tonotopic mapping for the localization of the primary auditory cortex (Da Costa et al., 2015; 2018).

The inferior frontal cortex contributes to the cognitive evaluation of emotional cues of verbal and non-verbal vocalizations and its modulation by attention (review Frühholz and Grandjean, 2013). The putatively homologous area in non-human primates, the lateral prefrontal cortex, receives relatively sparse afferents from AMY; this contrasts with the strong, bidirectional connections, which AMY has with the orbitofrontal and medial prefrontal areas (Ghashghaei and Barbas, 2002; Barbas, 2007; Barbas et al., 2011).

In summary, the early-stage auditory areas, VA and AMY constitute a core network for the processing of emotional vocalizations (Frühholz et al., 2016), which is fostered by strong connections, as demonstrated in hodological studies in non-human primates and in man. The primary auditory cortex and AMY receive both direct auditory input from the medial geniculate nucleus (Shinonaga et al., 1994). A complex pattern of cortico-cortical connections links the primary and non-primary early-stage areas and the adjacent superior temporal convexity (Cammoun et al., 2015). The extended auditory region is interconnected with AMY (Price and Amaral, 1981; Yukie, 2002; Reser et al., 2009). On the basis of this complex architecture, and notably a dual auditory input via the primary auditory cortex and via AMY, it can be argued that the processing of emotional vocalizations may differ between the early-stage auditory areas, VA and AMY, possibly with different selectivity in respect to that of other emotional sounds.

Here, we made use of the high spatial resolution of ultra-high field fMRI at 7 Tesla to investigate the representation of human vocalizations *vs*. other environmental sounds, and their modulation by emotional valence within early-stage auditory areas, VA and AMY. Based on previous findings, we expected (1) the AMY to process emotional valence both for vocalizations and non-vocalizations; (2) specific auditory belt areas to encode specifically human vocalizations or emotional valence but not emotional valence of human vocalizations only; and (3) VA and/or AMY to process emotional valence selectively for human vocalizations. These hypotheses were tested by comparing the BOLD responses within the above regions of interest to human vocalizations and to other environmental sounds with positive, neutral or negative emotional valence using various repeated measures ANOVA. In addition we explored functional coupling between individual early-stage auditory areas, VA and AMY, expecting to find a signatures of the dual auditory input via the primary auditory cortex and the amygdala.

## Materials and Methods

### *Subjects*

Eleven subjects (7 female, 8 right-handed, mean age 25.3 ± 4.27 years) participated in the study after giving written, informed consent. None of the participants reported hearing deficits or history of neurological or psychiatric illness. Hearing thresholds and mental states were measured prior to testing. All participants were native speakers of French, without musical training. All experimental procedures were approved by the Ethics Committee of the Canton de Vaud. The dataset of one subject was discarded due to data acquisition problems (ghosting), and data from the remaining ten subjects were used in the following analysis.

Participants provided informative health status and then completed five questionnaires: the Edinburgh Handedness Inventory (Oldfield, 1971), the Hospital Anxiety and Depression (HAD) scale (Zigmond, A. S. and Snaith, R. P., 1983), the Big Five Inventory (Plaisant et al., 2010), and a musical aptitude questionnaire. The participants were representative of the general population. The results of the Big-Five Inventory showed that the N (Neuroticism) score had greater between-subject variability than the other scores. The scores for factors A (Agreeableness), C (Conscientiousness) and E (Extraversion) in our subject sample were higher than in the sample from (Plaisant et al., 2010), whereas the N score was smaller and had greater variability. The distribution of the O (Openness) score of the current study was similar to that of (Plaisant et al., 2010). Concerning the HAD scale, the between-subject variability was greater for the anxiety score compared to the depression score. Despite this score variability, no subjects were excluded based on these results.

*Experimental design and statistical analysis*

The *experimental design* included a single fMRI session (~ 55-60 min in total) during which participants listened passively (i.e., without performing a task) to human vocalizations or other environmental sounds with positive, neutral or negative emotional valence while fixating on a red cross on a black background. The baseline condition was resting silently with the same fixation. The fMRI session was followed by a rating of the emotional valence of the 66 stimuli used in the experiment. A debriefing was then performed outside the MRI scanner.

*Auditory stimuli* were presented in blocks of eleven different sounds from the same category (human vocalizations or other environmental sounds) and with the same emotional valence (positive, neutral or negative). On the whole 6 stimulus conditions were presented: i) human vocalizations with neutral valence (vowels or consonant-vowels without significance); ii) human vocalizations with positive valence (e. g. baby or adult laughing, erotic vocalization by man or woman); iii) human vocalizations with negative valence (e. g. frightened scream, vomiting, brawl); iv) non vocalizations with neutral valence (e. g. running car engine, wind blowing, train); v) non-vocalizations with positive valence (e. g. applause, opening beer can and pouring into the glass; river); and vi) non-vocalizations with negative valence (e. g. ticking and exploding bomb; tire skids, breaking glass). Each subject listened to three runs, in which blocks and their sequence order were pseudo-randomized. Each fMRI run began with a 30-s silent rest condition, followed by 14 blocks, each of which lasted 30 s (22 s of sounds + 8 s of silence), followed again by a 30-s silent rest condition (total of 8min). Sounds were presented using MATLAB (R2015b, The MathWorks, Inc., Natick, Massachusetts, United States) and the Psychophysics Toolbox (www.psychtoolbox.org). Stimuli were delivered binaurally at 80 ± 8 dB SPL via MRI-compatible headphones (SensiMetrics S14, SensiMetrics, USA), following prior filtering with the SensiMetrics filters to obtain a flat frequency transmission. The auditory stimuli used in this experiment were the same as in (Aeschlimann et al., 2008), who showed in their study that human vocalization are a separate category within the environmental sounds. In this battery, 66 different emotional sound files of 2 s were selected and equally distributed in the following six categories: Human Vocalizations Positive (HVP), Human Vocalizations Neutral (HV0), Human Vocalizations Negative (HVN), Non-Vocalizations Positive (NVP), Non-Vocalizations Neutral (NV0), and Non-Vocalizations Negative (NVN). Categories were controlled for their acoustic characteristics: the percentage of points showing a significant difference between the mean spectrogram of two different sound categories was calculated and maintained below 1% to avoid acoustic differences between the six categories of sound, as in (De Meo et al., 2015). All the sounds were also tested using PRAAT software (http://www.fon.hum.uva.nl/praat/) and homemade MATLAB scripts to determine their mean fundamental frequency, mean intensity, harmonics to noise ratio, power, center of gravity, mean Wiener entropy and spectral structure variation (Reddy et al., 2009). Two-way repeated measures ANOVA with the factors Vocalization (Human-Vocalizations, Non-Vocalizations) x Valence (Positive, Neutral, Negative) were performed to compare the effect of each acoustic feature on the sound categories. We found a main effect of Vocalizations [F(1,64) = 18.68, p = 0.0015], a main effect of Valence [F(2,63) = 21.14, p = 1.17E-5] and an interaction Vocalizations x Valence [F(2,63) = 8.28, p = 0.002] on the mean Wiener entropy. We found a main effect of Valence [F(2,63) = 10.51, p = 0.0007] on the center of gravity. There was a main effect of Vocalizations [F(1,64) = 134.23, p = 4.06E-7], a main effect of Valence [F(2,63) = 69,61, p = 9.78E-10] and an interaction of Vocalizations x Valence [F(2,63) = 17.91, p = 3.48E-5] on the harmonics-to-noise ratio. Finally, there was an interaction of Vocalizations x Valence on the mean intensity [F(2,63) = 12.47, p = 0.0003] and on the power [F(2,63) = 14.77, p = 0.0001].

The post-acquisition rating of the emotional valence of each stimulus was performed while the subject was still lying in the scanner (without any sequence running) to minimize emotional bias and to match at best the experimental conditions. Stimuli were presented in random order; after each sound presentation, the subject was instructed to judge the valence of the sound with a linear visual 7-point scale (1 being very pleasant and 7 being very unpleasant) and to give their answer orally within a 5 s silent gap.

*Tonotopic mapping* was achieved by presenting each subject pure tones (88 – 8000 Hz, in half octave steps) in ordered progressions for 2 s, as described previously (Da Costa et al., 2011; 2013; 2015; 2018). A cycle was composed of 28 s of tone presentation (14 frequencies x 2 s) followed by a 12-s silent pause. A single fMRI run consisted of 12 identical cycles for a total duration of 8 min. Each subject listened to two runs, either with ascending or descending frequency progressions. One subject (the pilot of the study) had a different mapping paradigm (but comparable tonotopic maps), with seven tones (88 – 8000 Hz, in octave steps) and 12 cycles (composed of 14 s of tone presentation and 14 s of silent pause).

*The identification of the regions of interest (ROIs)* was performed as follows. First, the identification of the *early-stage auditory areas* was carried out as described previously (Da Costa et al., 2015; 2018). Briefly, individual tonotopic mappings were used to identify in each subject the primary and non-primary areas, which were designated as the primary auditory areas (A1 and R), the lateral belt areas (L1, L2, L3 and L4) and medial belt areas (M1, M2, M3 and M4). These individually defined areas were used as ROIs for the analysis of neural activity (Fig. 1A, orange box, Table 1), and their respective coordinates were in accordance with previously published values (Viceic et al., 2006; van der Zwaag et al., 2011; Da Costa et al., 2015; 2018). Second, VA was identified by the contrast

‘Human-Vocalizations *vs.* Non-Vocalizations’ (p = 0.001, uncorrected). All significant voxels on the posterior part of the superior temporal gyrus and sulcus were considered as part of VA. The Talairach coordinates of VA defined in this way were well within the general regions of VA as described by Pernet and colleagues (2015). For some subjects, a few voxels were located within the lateral early-stage auditory areas; they were not considered as belonging to VA but to the respective lateral early-stage auditory areas in further analysis. Third, AMY was identified in each subject on the anatomical images using BrainVoyager (BrainVoyager QX v2.8, Brain Innovation, Maastricht, Netherlands) drawing tools. Given that the same sounds are used for the definition of VA and for the following statistical analysis, the results for VA region are only descriptive, except for the post-hoc correlation analysis.

**Table 1: Mean Talairach coordinates (center of gravity) of all ROIs and mean areas. STD = standard deviation**

| ROI | X | ± | STD(X) | Y | ± | STD(Y) | Z | ± | STD(Z) | Area | ± | STD(Area) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Left hemisphere*** | | | | | | | | | | | | |
| **Amy** | -21.79 | ± | 4.32 | -4.62 | ± | 3.06 | -14.41 | ± | 3.42 | 1240 | ± | 218.96 |
| **A1** | -41.85 | ± | 4.77 | -25.63 | ± | 4.82 | 10.71 | ± | 3.87 | 781.1 | ± | 113.52 |
| **R** | -38.68 | ± | 4.18 | -20.58 | ± | 4.77 | 10.04 | ± | 3.86 | 736.2 | ± | 147.44 |
| **L1** | -53.67 | ± | 5.02 | -35.60 | ± | 8.63 | 16.97 | ± | 8.19 | 2204 | ± | 491.04 |
| **L2** | -54.40 | ± | 5.19 | -18.26 | ± | 5.00 | 9.22 | ± | 4.00 | 794.2 | ± | 89.6 |
| **L3** | -48.86 | ± | 5.46 | -8.73 | ± | 5.80 | 5.23 | ± | 3.51 | 1193 | ± | 286.6 |
| **L4** | -42.96 | ± | 4.16 | -0.63 | ± | 9.85 | -7.21 | ± | 6.98 | 2113 | ± | 305.24 |
| **M1** | -44.50 | ± | 6.44 | -35.80 | ± | 5.31 | 20.71 | ± | 7.75 | 1723 | ± | 352.72 |
| **M2** | -32.94 | ± | 2.71 | -31.42 | ± | 3.17 | 16.75 | ± | 2.87 | 254.7 | ± | 65.1 |
| **M3** | -30.05 | ± | 1.89 | -27.67 | ± | 3.15 | 16.89 | ± | 3.46 | 182.4 | ± | 55.08 |
| **M4** | -34.54 | ± | 3.19 | -11.19 | ± | 9.84 | -2.39 | ± | 10.71 | 1698 | ± | 237.5 |
| **VA** | -53.30 | ± | 6.21 | -30.36 | ± | 5.72 | 6.13 | ± | 4.43 | 435.4 | ± | 202.36 |
| ***Right hemisphere*** | | | | | | | | | | | | |
| **Amy** | 20.27 | ± | 4.36 | -5.05 | ± | 2.84 | -14.57 | ± | 3.42 | 1259 | ± | 180.96 |
| **A1** | 43.24 | ± | 4.83 | -26.09 | ± | 5.00 | 11.79 | ± | 3.71 | 678.3 | ± | 110.16 |
| **R** | 40.24 | ± | 4.24 | -20.24 | ± | 4.96 | 8.71 | ± | 4.29 | 745.9 | ± | 141.28 |
| **L1** | 55.52 | ± | 4.80 | -31.55 | ± | 5.86 | 19.18 | ± | 9.55 | 1903 | ± | 436.44 |
| **L2** | 56.71 | ± | 4.23 | -21.04 | ± | 6.27 | 9.50 | ± | 4.49 | 922.5 | ± | 195.7 |
| **L3** | 52.15 | ± | 5.07 | -9.87 | ± | 5.78 | 3.67 | ± | 3.56 | 1003 | ± | 174.8 |
| **L4** | 43.62 | ± | 4.71 | 0.31 | ± | 10.10 | -7.29 | ± | 6.39 | 2380 | ± | 343.2 |
| **M1** | 44.73 | ± | 6.25 | -32.77 | ± | 5.04 | 24.99 | ± | 8.42 | 1477 | ± | 144.36 |
| **M2** | 33.25 | ± | 3.11 | -30.77 | ± | 4.48 | 17.78 | ± | 4.36 | 236.1 | ± | 35.3 |
| **M3** | 31.47 | ± | 2.30 | -26.75 | ± | 3.97 | 16.23 | ± | 4.21 | 199.3 | ± | 43.56 |
| **M4** | 34.95 | ± | 2.92 | -10.74 | ± | 10.96 | -3.30 | ± | 10.32 | 1844 | ± | 279.36 |
| **VA** | 48.79 | ± | 7.60 | -31.39 | ± | 7.37 | 5.46 | ± | 4.98 | 592.3 | ± | 222.56 |

*MRI data acquisition* was performed on a 7-Tesla MRI scanner (Siemens MAGNETOM scanner, Siemens Medical Solutions) with an 8-channel head rf-coil (RAPID Biomedical). In order to acquire high spatial resolution datasets, a sinusoidal 2D-EPI sequence with 1.5 mm isotropic voxels was used for the functional acquisition (1.5 x 1.5 mm in-plane resolution, slice thickness = 1.5 mm, TR = 2000 ms, TE = 25 ms, flip angle = 70°, slice gap = 0 mm, matrix size = 146 x 146, field of view = 219 x 219, with 43 oblique slices centred on the superior temporal plane along the lateral sulcus, with a total coverage of ~ 65 mm and covering the full extent of the superior and medial temporal sulci until the entorhinal cortex). T1-weighted high-resolution 3D anatomical images were acquired with a MP2RAGE sequence (resolution = 1 x 1 x 1 $mm^3$, TR = 5500 ms, TE = 1.87 ms, TI1/TI2 = 750/2350 ms, slice gap = 0 mm, matrix size = 256 x 240, field of view = 256 x 240 (Marques et al., 2010).

The processing of emotional stimuli in AMY has been shown to modulate autonomic reactions, via specific outputs to brain stem nuclei (Janak and Tye, 2015). Emotional stimuli of different valence can thus induce distinct changes in heart rate or breathing and introduce a bias to image analysis. We accounted for this in our study and recorded during the experiment, pulse oximetry and respiration, using a plethysmograph and respiratory belt provided from the MRI scanner vendor. In total, each subject had an imaging session of 54 min with five functional runs: three runs of the auditory emotional experiment and two runs of the tonotopic mapping experiment, which were used for the definition of the ROIs within the superior temporal plane.

*The MRI analysis* included the following steps. Preprocessing steps included scan time correction (only for the auditory emotional runs), temporal filtering, motion correction, segmentation and normalization into the Talairach space and were performed with BrainVoyager. These preprocessing steps were common to all fMRI acquisitions, then depending on the purpose of the fMRI run, the datasets were processed differently. In order to define early-stage auditory areas, a linear cross-correlation analysis was computed for each tonotopic mapping dataset, and the resulting correlation maps were averaged together (ascending and descending correlation map) to define the best frequency value for each voxel according to the cycle order (Da Costa et al., 2011; 2013; 2015; 2018). These analyses were performed at the single subject level in the volumetric space, and the resulting maps were projected onto the cortical surface meshes, where the individual early-stage auditory areas and VA were defined and projected back to the volumetric space (see the paragraph on “the identification of the regions of interest (ROIs)” for more details). Then, we performed a random effects (RFX) group analysis on the auditory emotional runs, with movement and respiration parameters as regressors, and tested for the contrast ‘Sounds *vs.* Silence’ with an FDR correction at q

$< 0.05$ ($p < 0.05$, Figure 1.B). This GLM analysis was used to verify that specific regions of the brain were activated by our paradigm and particularly that our ROIs were activated by the emotional sounds. The remaining analysis focused only on the BOLD responses extracted from these ROIs, as explained in the next paragraphs.

*The BOLD time course extraction and processing* were performed as follows. Functional individual time courses for each ROI were extracted in the 3D volume space using BrainVoyager and imported into MATLAB. Each time course was normalized by its mean signal, separated according to the sound category, and averaged (1) spatially within each ROI, (2) temporally over blocks and runs, and (3) across the ten subjects, resulting in a time course of 15 time points for each ROI and category. A time-point-by-time-point 2-Way repeated measure ANOVA, 2 Vocalization (Human-Vocalizations, Non-Vocalizations) x 3 Valence (Positive, Neutral, Negative) was performed on the averaged BOLD time courses according to (Da Costa et al., 2015; 2018). Significant results were restricted temporally by only considering the p-values lower or equal to 0.05 for at least three consecutive time points. It is to be noted that the probability that three consecutive time-points be false positives is ($0.05 \times 0.05 \times 0.05$) = $1.25*10^{-4}$. Therefore, this constraint in time was considered as a valid correction for our analysis (see Da Costa et al., 2015 for more details). Finally, post hoc time-point-by-time-point paired t-tests were performed between each pair of sound categories.

*Physiological data* were processed with the TAPAS PhysIO toolbox (Kasper et al., 2017). The respiration recordings were used as regressors in the GLM model, whereas the cardiac recordings were processed with the same pipeline as the BOLD signal to obtain a pulse time course for each sound category. Heart rate, interbeat interval (time interval between two successive beats) time courses and heart rate variability were also extracted from these data. The heart rate variability was calculated using the root mean squared successive difference between the interbeat interval, normalized by the mean interbeat interval according to (Goedhart et al., 2007). These latter measures were used to evaluate the effect of the emotional content of the auditory stimuli on the cardiac rhythm.

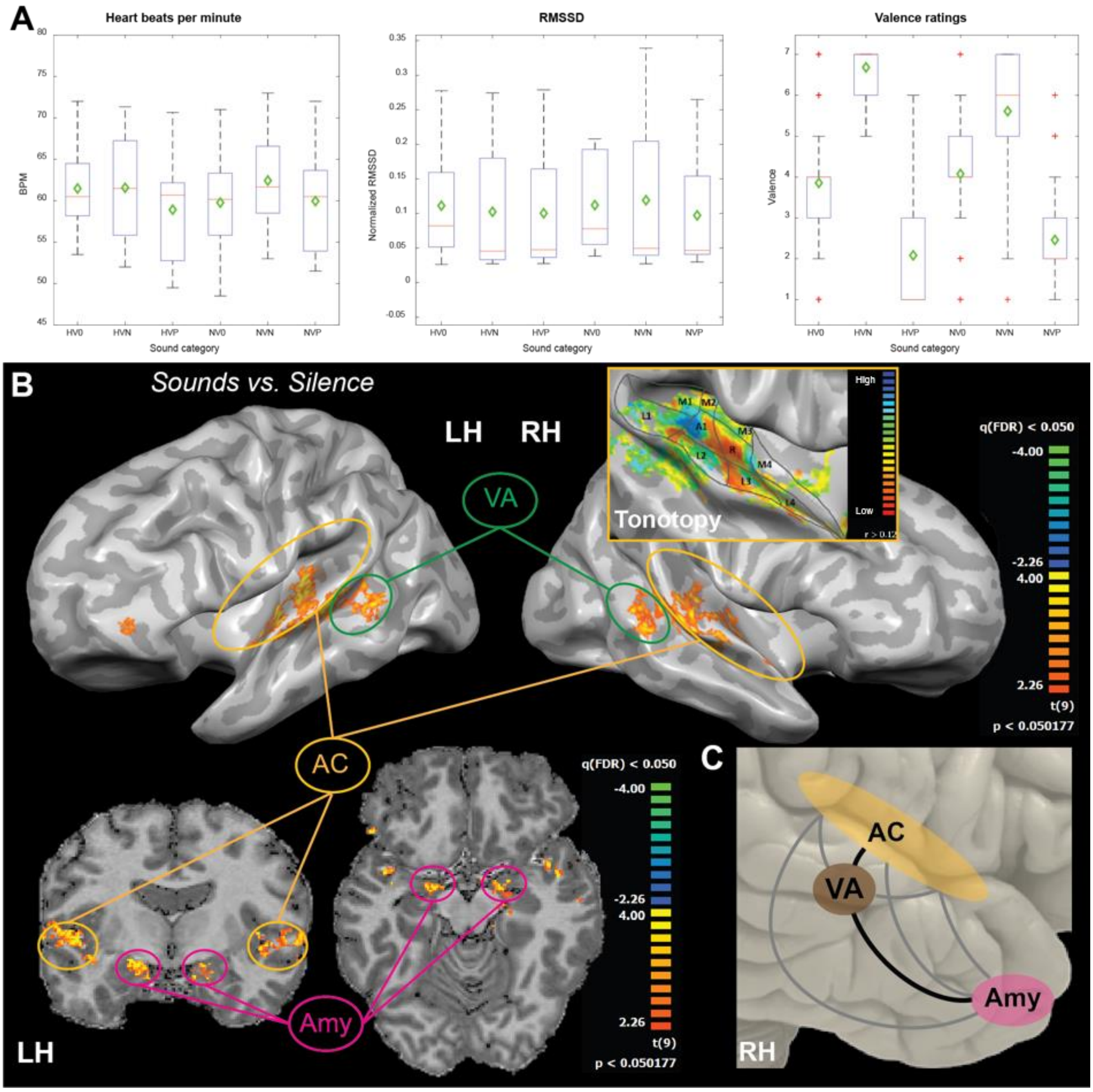

**Figure 1: A. Behavioral results.** Left panel: Heart rate (beats per minute). Middle panel: Root mean squared successive difference (RMSSD). Right panel: Valence ratings. The red line represents the median value, the box indicates the values between the 25th and the 75th percentiles, and the whiskers show the data below the 25th percentile or above the 75th percentile, not considered as outliers. The outliers are depicted with a plus symbol. The green diamond represents the mean value. Refer to the manuscript for the sound categories abbreviations. **B. GLM statistical maps and ROI definition.** The maps resulting from the contrast 'Sounds *vs.* Silence' are presented in the left panel on the volume in the coronal and transverse views and in the right panel on the surface (FDR corrected, $q < 0.05$). Orange box: tonotopic maps were projected into the individual right hemisphere surface of a representative subject ($r > 0.12$). The frequency-selective region was divided into 10 ROIs: A1, R, L1, L2, L3, L4, M1, M2, M3 and M4. **C. Hypothesis.** We expected (1) the AMY to process emotional valence both for vocalizations and non-vocalizations and to be highly correlated with VA and specific lateral belt auditory areas (L1, L2, and L3; grey lines); (2) specific auditory belt areas to encode specifically human vocalizations or emotional valence but not emotional valence of human vocalizations only and to be correlated to each other; and (3) VA and/or AMY to process emotional valence selectively for human vocalizations and being modulated by lateral belt auditory areas (L2; dark lines). RH: right hemisphere; LH: left hemisphere. Refer to the manuscript for the ROIs definition and abbreviations.

## Results

### *Modulation of behavioural measures by emotional and vocal content*

The average heart rate did not show any significant differences between sound categories (Fig. 1.A, left). Heart rate variability, represented by the normalized RMSSD (Root Mean Squared Successive Difference), did not show any significant difference between sound categories (**Erreur ! Source du renvoi introuvable.**.A, middle). The time courses of the pulses were submitted to a time-point-by-time-point 2-way repeated measures ANOVA 2 Vocalization (Human-Vocalizations, Non-Vocalizations) x 3 Valence (Positive, Neutral, Negative), which highlighted a main effect of vocalization.

Post-scanning ratings of the valence of the sound stimuli showed a bigger variance in the categories HVP and NVN compared to that in the four other sound categories (**Erreur ! Source du renvoi introuvable.**.A, right). A 2-way repeated measures ANOVA 2 Vocalization (Human-Vocalizations, Non-Vocalizations) x 3 Valence (Positive, Neutral, Negative) on the valence ratings revealed a main effect of valence [F(2, 653) = 532.29, p = 7.38E-138], no effect of Vocalization [F(1, 653) = 2.68, p = 0.1], and an interaction of Vocalization x Valence [F(2, 653) = 22.31, p = 4.23E-10]. As indicated by post-hoc t-tests, the latter was driven by the difference between vocalizations vs non-vocalizations, which was significant for negative, but not neutral or positive valence.

### *Emotional modulation of neural activity elicited by human vocalizations and by other environmental sounds*

The RFX GLM analysis with the contrast 'Sounds *vs.* Silence' ($p < 0.005$, $q(FDR) < 0.05$) resulted in a strong bilateral activation on the supratemporal plane, the posterior part of the superior temporal sulcus corresponding to VA and in the AMY (Fig. 1.B).

**Table 2: Talairach coordinates of the peaks of the activation clusters of *Sounds vs. Silence*. Talairach coordinates, t values, number of voxels and corresponding regions for the peaks of all the activation clusters resulting from the contrast Sounds vs. Silence in the RFX GLM analysis (FDR q < 0.05, cluster threshold of 100 voxels). Clusters were sorted according their size and corresponding region. HG: Heschl's Gyrus; IFG: Inferior Frontral Gyrus; PFC: Prefrontral Gyrus; STG: Superior Temporal Gyrus; ITG: Inferior Temporal Gyrus; MGB: Medial Geniculate Body of the thalamus; AMY: Amygdala.**

| Area | X | Y | Z | t value | #voxels | BA |
|---|---|---|---|---|---|---|
| ***Left hemisphere*** | | | | | | |
| **Transverse gyrus (HG)** | -48 | -19 | 10 | 33.75 | 21230 | 41 |
| **Triangular part IFG** | -45 | 16 | 16 | 10.05 | 1282 | 45 |
| | -42 | 19 | 10 | 9.33 | 771 | |
| **Dorsolateral PFC** | -47 | 42 | 7 | 9.73 | 392 | 46 |
| | -51 | 29 | 13 | 7.41 | 247 | |
| **Parahippocampal gyrus** | -18 | -11 | -13 | 9.23 | 386 | 54 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Orbital part IFG** | -28 | 27 | -2 | 9.53 | 354 | 47 |
| **Insula** | -24 | 23 | 12 | 7.11 | 307 | 13 |
| **STG** | -52 | -51 | 12 | 7.74 | 164 | 22 |
| **Cerebellum** | -41 | -58 | -23 | 7.77 | 154 | - |
| | -23 | -29 | -22 | 8.72 | 145 | - |
| **ITG** | -54 | -57 | -15 | 7.63 | 150 | 20 |
| **MGB** | -16 | -26 | -5 | 8.46 | 140 | - |
| **Cuneus (V1)** | -11 | -90 | -8 | 7.11 | 137 | 17 |
| ***Right hemisphere*** | | | | | | |
| **Transverse gyrus (HG)** | 54 | -16 | 5 | 31.22 | 24633 | 41 |
| **Dorsolateral PFC** | 49 | 38 | 2 | 10.42 | 922 | 46 |
| **AMY** | 16 | -5 | -20 | 11.30 | 715 | 53 |
| **Opercular part IFG** | 45 | 16 | 21 | 11.87 | 615 | 44 |
| | 54 | 16 | 30 | 8.51 | 184 | |
| | 36 | 30 | 0 | 6.76 | 159 | |
| | 31 | 29 | 0 | 5.78 | 121 | |
| **Dorsolateral PFC** | 50 | 34 | 19 | 8.18 | 213 | 9 |
| **Inferior Colliculus** | 6 | -32 | -2 | 11.95 | 144 | - |
| **Secondary visual cortex** | 7 | -88 | -15 | 9.12 | 135 | 18 |
| **Thalamus** | 11 | -14 | 6 | 8.29 | 111 | - |

Two-way repeated measure ANOVA on the BOLD responses with factors Vocalization (Human-Vocalizations, Non-Vocalizations) and Valence (Positive, Neutral, Negative) revealed a significant main effect of Vocalization bilaterally in VA and L2, as well as in the left L1 and right L3 (Figure 2.A). A main effect of Valence was significant bilaterally in VA and L3, as well as in the left L2, right L4 and AMY. The interaction Vocalization x Valence was significant bilaterally in VA, as well as in right A1. To investigate this significant interaction, we performed post-hoc one-way ANOVAs for the factor valence on the categories of human vocalizations and non-vocalizations separately, as well as post-hoc one-way ANOVAs for the factor vocalization on the categories of positive, negative and neutral sounds separately (Figure 2.B). The analysis of the effect of emotional valence on human vocalizations (HVP, HV0, HVN) using one-way ANOVA yielded a significant main effect of Valence bilaterally in VA. The analysis of the effect of emotional valence on non-vocalizations (NVP, NV0, NVN) using one-way ANOVA did not yield any significant main effect. Post-hoc t-tests reveal that the main effect of Valence on vocalizations in VA was driven by a significant difference between 'HVP and HVN' and 'HV0 and 'HVN'. The interaction effect found in right A1 was driven by a main effect of vocalization only for the positive stimuli and not for the neutral nor negative stimuli. A post hoc three-way repeated measure ANOVA with factors Vocalization (Human-Vocalizations, Non-Vocalizations), Valence (Positive, Neutral, Negative) and Time (every 15 time bins of the block) revealed a significant main effect for Vocalization in bilateral VA, in left

L1 and L2, and right L3; a significant main effect for Valence in right L3 and bilateral VA; a significant main effect for Time in all bilateral ROIs; a significant interaction Vocalization X Time bilaterally in R, L2, L3, VA, in left L1 and M3, and right M1 and M2; a significant interaction Valence X Time bilaterally in L2, L3, and VA; and a significant interaction Vocalization X Valence X Time in bilateral L2, L3, VA, right AMY and left A1 and R. There was no significant interaction Vocalization X Valence when Time was considered as a factor (results not displayed). The main effect for Time in all bilateral ROIs supported our main hypothesis that the BOLD evolved differently for each stimuli along the blocks, therefore our interpretations are focusing on the results from the two-way ANOVA.

The BOLD time courses for the ROIs with a significant effect in the ANOVA are presented in Figure 3. VA responded preferentially to human-vocalizations of neutral and positive valence, whereas the STG responded preferentially to neutral sounds. The AMY shows a stronger response for the positive emotional sounds. We observed that the processing of the various valences occurs at the beginning of the time course, regardless of the ROI. Moreover, we noticed a habituation effect for all sound categories in all ROIs, with the strongest one (i.e. longest plateau) observed in VA for the vocalizations.

*Correlation of BOLD signal between regions of interest*

To investigate the coupling between ROIs with significant effects of the 2-way repeated measure ANOVA of the BOLD responses, we performed post hoc correlations between the BOLD signals of the AMY, VA, L1, L2 and L3 for Human Vocalizations and for non-vocalizations (Fig. 4.A). Significant correlations ($p < 0.01$; with an adjusted $R^2 > 0.6$; see Table 2 and 3) were found among the three belt areas (L1, L2, L3) and VA; their strength varied as a function of stimulus category and partially valence and between the hemispheres. Strikingly, the correlations between VA and the belt areas were stronger for Vocalizations than for Non-vocalizations. The right AMY was correlated with the right VA and right L1, L2 and L3 during the presentation of HVP (but not during Non-vocalizations nor during HVN or HV0); this effect was not found in the left hemisphere.

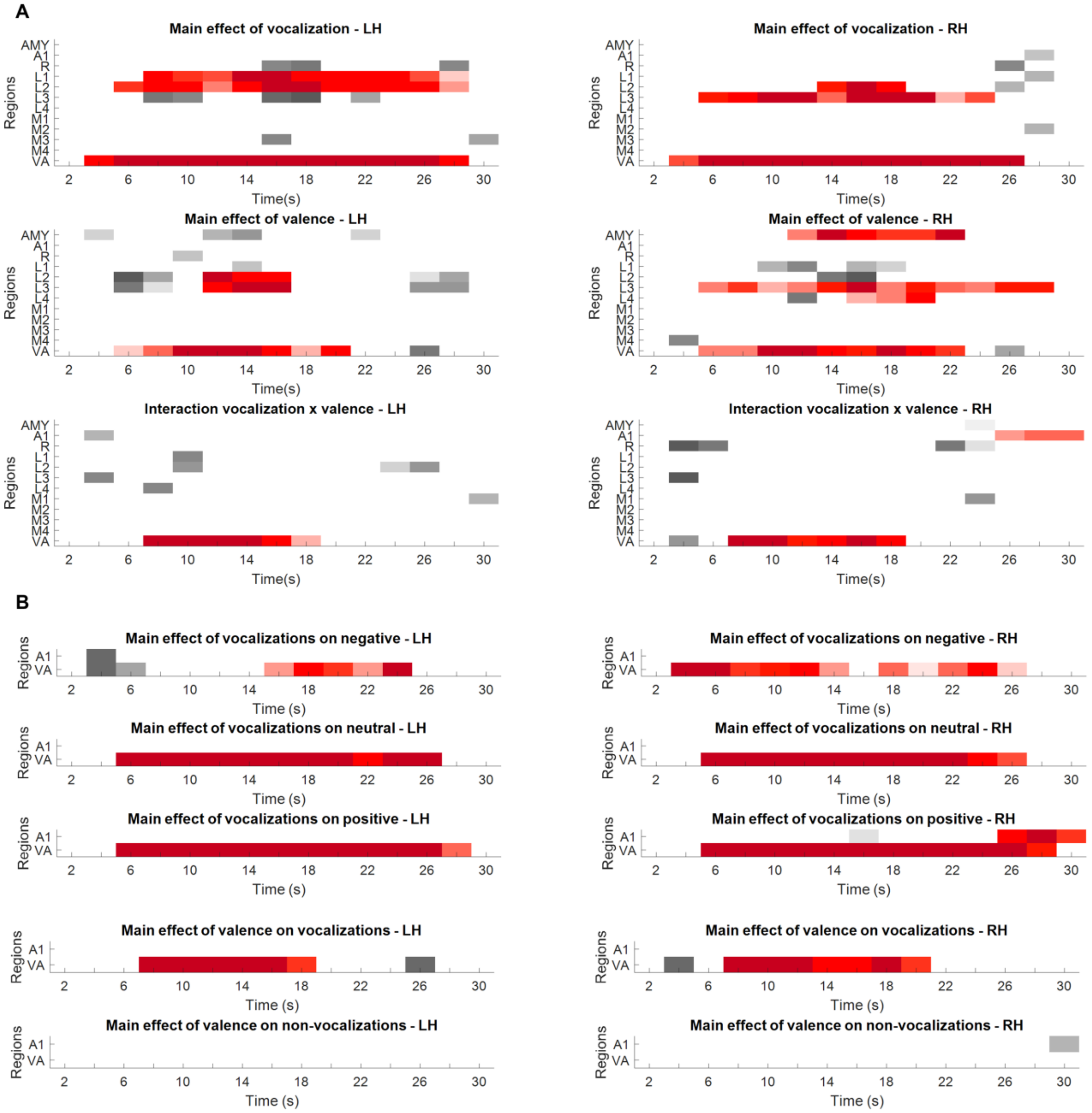

**Figure 2: Statistical results. A. Two-way ANOVA 2 Vocalization (Human Vocalizations, Non-Vocalizations) x 3 Valence (Positive, Neutral, Negative) on the BOLD signal.** The results of the ANOVA Vocalization x Valence, with the ROIs represented on the y-axis, the time points on the x-axis and the color bar indicating different statistical thresholds. The red colors indicate a p-value lower or equal to 0.05 for at least three consecutive time points. Upper panel: Main effect of Vocalization. Middle panel: Main effect of Valence. Lower panel: Interaction Vocalization x Valence. **B. Separate one-way ANOVA for human vocalizations and non-vocalizations.** Upper panel: Main effect of Valence on Human Vocalization categories (HVP, HV0, HVN). Lower panel: Main effect of Valence on Non-Vocalization categories (NVP, NV0, NVN). RH: right hemisphere; LH: left hemisphere. Refer to the manuscript for the ROIs definition and abbreviations.

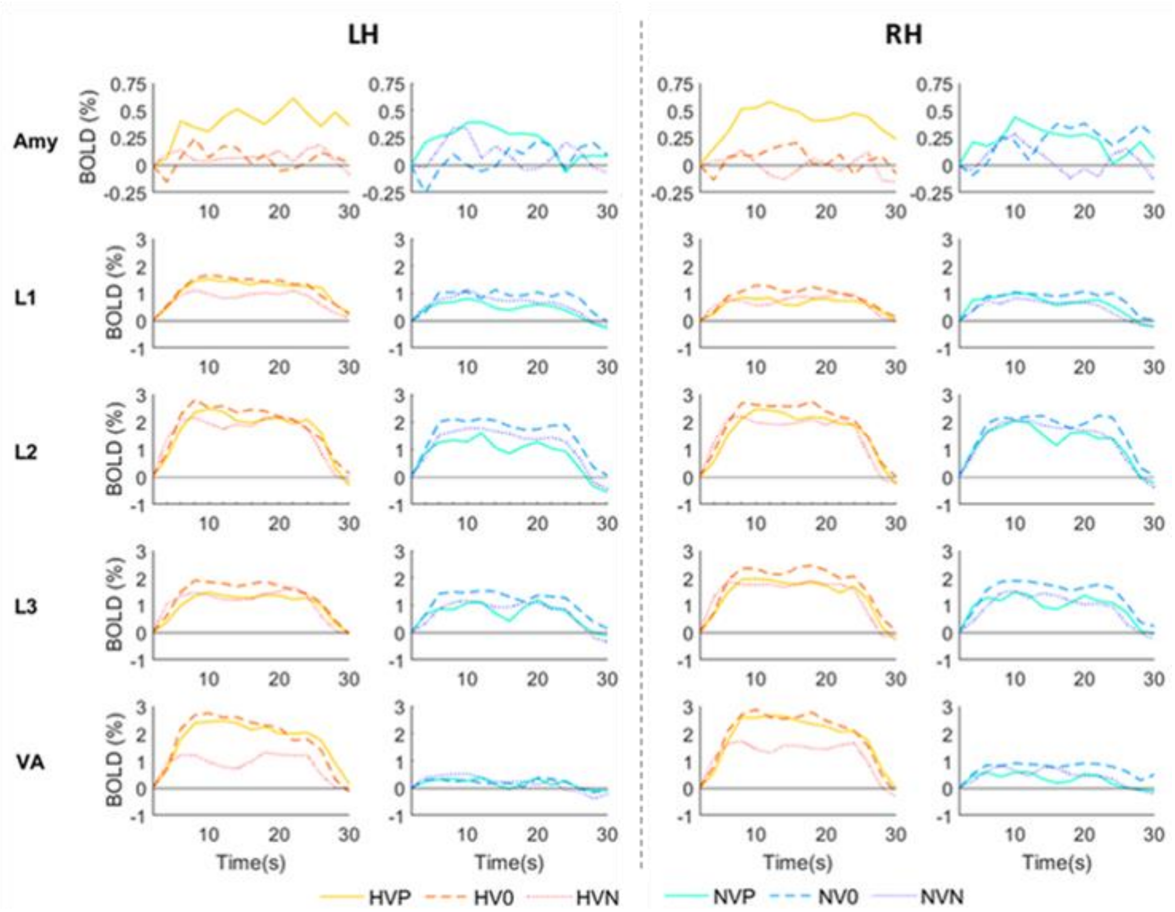

**Figure 3: BOLD time courses for significant ROIs: AMY, L1, L2, L3 and VA.** BOLD time courses for the left and right hemisphere are in the left and right part, respectively. For each hemisphere, the left panel depicts the Human Vocalization categories [HVP (solid line), HV0 (dashed line), HVN (dotted line)] in warm colors and the right panel the Non-Vocalization categories [NVP (solid line), NV0 (dashed line), NVN (dotted line)] in cold colors. RH: right hemisphere; LH: left hemisphere. Refer to the manuscript for the ROIs and sound categories abbreviations.

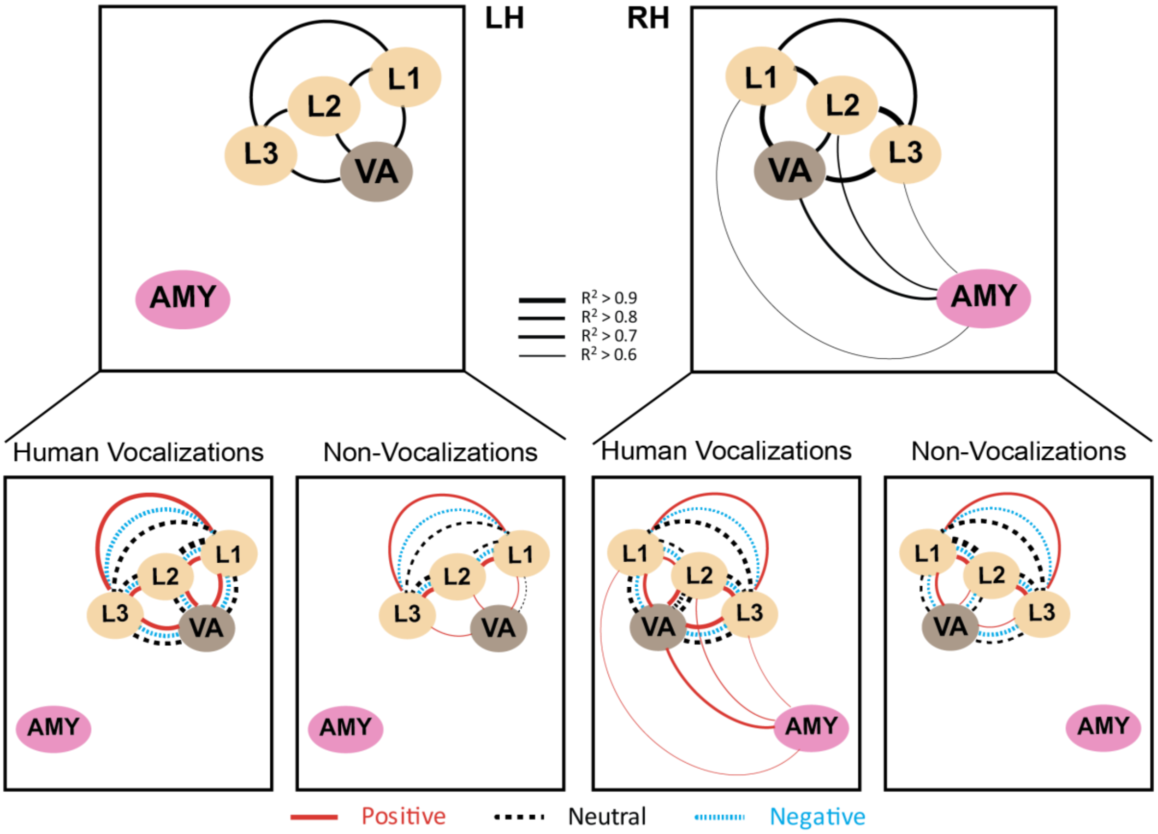

**Figure 4: BOLD correlations.** Correlations between the BOLD time courses of the ROIs L1, L2, L3, VA and AMY, in the left and right hemispheres separately. The thickness of the lines represents the strength of the correlation. Correlations with an adjusted-$R^2$ smaller than 0.6 are not represented. In the lower part of the figure, the correlations are separated for the Human Vocalizations categories [HVP (solid red line), HV0 (dashed black line), HVN (dotted blue line)] and for the Non-Vocalizations categories [NVP (solid red line), NV0 (dashed black line), NVN (dotted blue line)]. RH: right hemisphere; LH: left hemisphere. Refer to the manuscript for the ROIs definition and abbreviations.

**Table 3: BOLD correlations for the left hemisphere.** P-values and R-square for the BOLD correlations of the left ROIs (L1, L2, L3, VA and AMY) for the different sound categories (HV0, HVP, HVN, NV0, NVP and NVN). The results are only presented for correlations with p-values lower than 0.01 and R-square greater than 0.6. The correlations that do not meet this criterion are hatched. No result in the shaded part of the table, as this is a symmetrical matrix. Refer to the manuscript for the ROIs definition and abbreviations.

| *Left hemisphere* | **L1** | **L2** | **L3** | **VA** | **AMY** |
|---|---|---|---|---|---|
| | **p-value ($R^2$)** | **p-value ($R^2$)** | **p-value ($R^2$)** | **p-value ($R^2$)** | **p-value ($R^2$)** |
| **HV0** | | | | | |
| **L1** | | 1.56E-09 (0.94) | 8.36E-11 (09.96) | 2.08E-09 (0.94) | ▨ |
| **L2** | | | 1.56E-12 (0.98) | 5.72E-12 (0.98) | ▨ |
| **L3** | | | | 4.86E-11 (0.97) | ▨ |
| **VA** | | | | | ▨ |
| **AMY** | | | | | |
| **HVN** | | | | | |
| **L1** | | 2.38E-09 (0.94) | 2.20E-09 (0.94) | 3.73E-07 (0.87) | ▨ |
| **L2** | | | 6.73E-12 (0.98) | 1.85E-08 (0.92) | ▨ |
| **L3** | | | | 5.38E-09 (0.93) | ▨ |
| **VA** | | | | | ▨ |
| **AMY** | | | | | |
| **HVP** | | | | | |
| **L1** | | 5.65E-09 (0.93) | 5.05E-11 (0.97) | 5.42E-13 (0.98) | ▨ |
| **L2** | | | 1.14E-12 (0.98) | 3.85E-10 (0.96) | ▨ |
| **L3** | | | | 1.73E-11 (0.97) | ▨ |
| **VA** | | | | | ▨ |
| **AMY** | | | | | |
| **NV0** | | | | | |
| **L1** | | 6.99E-08 (0.89) | 6.74E-08 (0.89) | ▨ | ▨ |
| **L2** | | | 3.07E-13 (0.99) | ▨ | ▨ |
| **L3** | | | | ▨ | ▨ |
| **VA** | | | | | ▨ |
| **AMY** | | | | | |
| **NVN** | | | | | |
| **L1** | | 3.04E-09 (0.94) | 4.24E-08 (0.91) | 2.73E-05 (0.75) | ▨ |
| **L2** | | | 3.20E-10 (0.96) | ▨ | ▨ |
| **L3** | | | | ▨ | ▨ |
| **VA** | | | | | ▨ |
| **AMY** | | | | | |
| **NVP** | | | | | |
| **L1** | | 1.97E-10 (0.96) | 9.74E-08 (0.89) | 5.06E-05 (0.73) | ▨ |
| **L2** | | | 5.24E-08 (0.90) | 2.30E-05 (0.76) | ▨ |
| **L3** | | | | 9.46E-06 (0.79) | ▨ |
| **VA** | | | | | ▨ |
| **AMY** | | | | | |

**Table 4: BOLD correlations for the right hemisphere.** Same conventions as in Table 3

| *Right hemisphere* | L1 p-value ($R^2$) | L2 p-value ($R^2$) | L3 p-value ($R^2$) | VA p-value ($R^2$) | AMY p-value ($R^2$) |
|---|---|---|---|---|---|
| **HV0** | | | | | |
| **L1** | | 1.60E-10 (0.96) | 6.54E-10 (0.95) | 4.82E-11 (0.97) | ▨ |
| **L2** | | | 1.08E-12 (0.98) | 4.31E-15 (0.99) | ▨ |
| **L3** | | | | 3.14E-13 (0.98) | ▨ |
| **VA** | | | | | ▨ |
| **AMY** | | | | | |
| **HVN** | | | | | |
| **L1** | | 1.43E-06 (0.84) | 2.29E-06 (0.83) | 1.34E-07 (0.89) | ▨ |
| **L2** | | | 3.48E-13 (0.98) | 1.71E-10 (0.96) | ▨ |
| **L3** | | | | 3.01E-09 (0.94) | ▨ |
| **VA** | | | | | ▨ |
| **AMY** | | | | | |
| **HVP** | | | | | |
| **L1** | | 2.08E-07 (0.88) | 1.01E-07 (0.89) | 3.45E-07 (0.87) | 1.61E-04 (0.68) |
| **L2** | | | 9.18E-12 (0.97) | 4.48E-13 (0.98) | 2.96E-05 (0.75) |
| **L3** | | | | 1.76E-10 (0.96) | 1.48E-04 (0.68) |
| **VA** | | | | | 3.69E-06 (0.79) |
| **AMY** | | | | | |
| **NV0** | | | | | |
| **L1** | | 5.88E-10 (0.95) | 5.42E-10 (0.95) | 4.65E-07 (0.87) | ▨ |
| **L2** | | | 1.79E-12 (0.98) | 1.99E-06 (0.83) | ▨ |
| **L3** | | | | 2.70E-07 (0.88) | ▨ |
| **VA** | | | | | ▨ |
| **AMY** | | | | | |
| **NVN** | | | | | |
| **L1** | | 1.12E-08 (0.92) | 4.55E-07 (0.87) | 2.43E-06 (0.83) | ▨ |
| **L2** | | | 3.82E-11 (0.97) | 4.33E-07 (0.87) | ▨ |
| **L3** | | | | 3.98E-08 (0.91) | ▨ |
| **VA** | | | | | ▨ |
| **AMY** | | | | | |
| **NVP** | | | | | |
| **L1** | | 2.26E-08 (0.92) | 6.72E-08 (0.89) | 2.87E-07 (0.88) | ▨ |
| **L2** | | | 8.02E-10 (0.95) | 3.26E-05 (0.75) | ▨ |
| **L3** | | | | 7.76E-06 (0.79) | ▨ |
| **VA** | | | | | ▨ |
| **AMY** | | | | | |

[insert Figure 4 here]

## Discussion

Our results indicate that emotional valence modulates differentially neural activity that is elicited by human vocalizations vs. non-vocal environmental sounds within individual early-stage auditory areas, VA and AMY. Using 7T fMRI and 1.5 x 1.5 x 1.5 mm voxel size, without smoothing, rendered individual auditory areas accessible to investigation. Whereas emotional content modulates both vocalizations and other environmental sounds in early-stage auditory areas, it singles out vocalizations in VA. Both types of sounds are also modulated in AMY. The specificity profiles within these regions and the correlations in their activity suggest that VA shares emotional information both with early-stage auditory areas and with AMY.

### *Selectivity for emotional vocalizations*

Whereas emotional valence modulates neural activity elicited by different sensory modalities in AMY or by different sound categories in lateral belt areas, its modulation appears to be limited to a single category, human vocalizations, in VA. This area was initially identified by its selectivity for human vocalizations, including speaker's identity, over other environmental sounds or acoustically similar control stimuli such as scrambled voices or amplitude modulated noise (Belin et al., 2000, 2002; Warren et al., 2006; Latinus et al., 2013; Zäske et al., 2017). VA is located within the superior temporal sulcus and there is inter-individual variability as to its precise location (Pernet et al., 2015). VA neural activity elicited by human vocalizations was shown to be modulated by emotional intensity of happy or angry intonation (Ethofer et al., 2006b), of laughter (Lavan et al., 2017) and of positive or negative valence of non-verbal vocalizations (Bestelmeyer et al., 2017). Our results confirm the selectivity of VA for human vocalizations over other environmental sounds and show that emotional modulation impacts the encoding of vocalizations but not of other environmental sounds. To our knowledge, none of the previous studies investigated this issue specifically, and none reported emotional modulation of environmental sounds other than vocalizations within VA. Further studies need to establish whether the neural populations, which encode emotional vocalizations do so in a categorical way (positive vs. neutral vs. negative) or on a continuous scale.

### *Selectivity for emotions and for vocalizations within early-stage auditory areas*

Our results indicate that specific lateral belt areas are selective for vocalizations over other environmental sounds and/or modulated by emotional valence but that emotional modulation is not limited to a specific stimulus category.  Within the left hemisphere areas, L1 and L2, which are located postero-laterally on the planum temporale, are selective for vocalizations, whereas L2 and L3, which are located laterally on the planum temporale and Heschl's gyrus, are selective for emotional valence. Within the right hemisphere L2 and L3 are selective for vocalizations and L3 and L4 for emotional valence. Voice selectivity within the planum temporale has been documented in previous studies, showing stronger activation to vocal than to non-vocal sounds (Belin et al., 2000) and participating in spectrotemporal analysis of vocalizations, a processing step which is believed to precede speaker identification in the superior temporal sulcus (Warren et al., 2006). Modulation by emotional valence was reported in a region located posterolaterally to the PAC, where emotional vocalizations yielded stronger activations than neutral voices (Wildgruber et al., 2004a; Ethofer et al., 2006b, 2012; Leitman et al., 2010; Bestelmeyer et al., 2017) or distinct spatial response patterns to different emotional categories (Ethofer et al., 2009b). This region on the posterolateral supratemporal plane is very likely part of the belt or parabelt areas, as suggested by its connectivity pattern. Diffusion Spectrum MRI (DSI) and post-mortem tracing studies have shown that this part of the auditory cortex i) is interconnected with the primary auditory cortex and with higher-order areas on the superior temporal gyrus (Cammoun et al., 2015); ii) receives monosynaptic callosal afferents from the fusiform gyrus (Di Virgilio and Clarke, 1997); and iii) has intrinsic connections that tend to be longer than those within the primary auditory cortex, but shorter than those of Broca's area (Tardif and Clarke, 2001; Tardif et al., 2007). This supratemporal region was included, together with a large part of VA, in the so-called "emotional voice area" (Ethofer et al., 2012) because of its responsiveness to emotionally modulated vocalizations; the specificity of the emotional effect for vocalizations vs. other sound categories has, however, not been investigated prior to our study. The high spatial resolution of the present study allowed us to show that the "emotional voice area" consists of two functionally distinct regions, the early-stage auditory areas, where the emotional content modulates neural activity elicited by vocalizations and by other environmental sounds, and VA, where it modulates responses to vocalizations only. The effect appears to be driven by a stronger response to neutral than positive or negative valence in early-stage areas and by positive valence in VA.

### *Modulation by emotional valence in the amygdala*

Emotional valence in AMY is encoded independently of stimulus category. Although several studies have shown that AMY plays an important role in processing emotions in non-verbal vocalizations (Phillips et al., 1998; Morris et al., 1999; Sander et al., 2003, 2007; Fecteau et al., 2007; Frühholz et al., 2014), modulation by emotional valence concerns other auditory categories, demonstrated here and in a previous study (Frühholz et al., 2014) or other sensory modalities (Baas et al., 2004; Costafreda et al., 2008; Ball et al., 2009). Our finding that the emotional effect was driven by stronger responses to positive than neutral and negative stimuli is consistent with previous studies using human vocalizations (Fecteau et al., 2007; Wiethoff et al., 2009) or other stimuli (O'Doherty et al., 2001; Anderson et al., 2003; Winston et al., 2005; Ball et al., 2007; Hurlemann et al., 2008; Sergerie et al., 2008; Costa et al., 2010). However, our results contrast with a recent report that screams and alarms, a priori negative stimuli, activate strongly AMY; the authors attribute this selectivity to the acoustic feature of roughness, which is shared by both types of stimuli (Arnal et al., 2015). Several other studies have highlighted the preference of AMY for negative emotional valence (Morris et al., 1998; Phillips et al., 1998, 2001; Wright et al., 2001). In our experimental paradigm modulation by emotional valence was stronger in AMY in the right hemisphere. Previous studies reported right (for laughing and crying sounds: (Sander et al., 2003) or left lateralization (for non-linguistic lateralizations: (Fecteau et al., 2007) or bilateral activation (Aubé et al., 2015). These divergent findings could be explained by the complex structure of AMY, which is composed of several nuclei (Amunts et al., 2005; Roy et al., 2009; Solano-Castiella et al., 2011). Imaging AMY remains difficult because of inhomogeneities in the local magnetic field (Labar et al., 2001), the proximity of large veins (Boubela et al., 2015) and the lateralization of AMY activation due to the phase-encoding polarity (Mathiak et al., 2012). These limitations prevented us from parcellating AMY accurately and exploring emotional encoding in specific sub-nuclei.

### ***Processing pathway for emotional vocalizations***

Our results speak in favour of a module dedicated to the processing of the emotional value of human vocalizations but not of other environmental sounds, which is part of or co-extensive with VA (Fig. 4B). This observation highlights three features of emotional processing. First, the neural mechanisms underlying this specificity involve most likely the combination of a category-specific input from the lateral belt areas and of emotion-specific input from AMY, as suggested by evidence from activation and connectivity patterns. Our results indicate that emotional information, which is encoded in VA, shares a preference for positive stimuli with AMY and for neutral stimuli with lateral belt areas. As reported in previous studies, the lateral part of the planum temporale processes temporo-spatial information pertaining to vocalizations and relays this information to VA, where higher-order analysis, including voice identification, is conducted (Belin et al., 2000; Warren et al., 2006). The corresponding interconnection between lateral belt areas and the region of the superior temporal sulcus, where VA is located, was demonstrated using DSI tract tracing (Cammoun et al., 2015). Furthermore, functional connections were described between VA and the supratemporal region (Pernet et al., 2015). Functional connectivity between AMY and VA, reported in an early study (Roy et al., 2009), were not confirmed in a later study, which proposed that the AMY-VA link passes via the prefrontal cortex (Pernet et al., 2015). The model of dual input to VA, from the lateral belt areas and from AMY, is consistent with the multi-stage concept for the processing of vocalizations and valence (Schirmer and Gunter, 2017); evidence from EEG studies suggests that vocalizations and valence are processed first independently, before being integrated in higher order auditory or frontal regions. At the neuronal level, we can only hypothesize as to the mechanisms that underlie the emotion-vocalization selectivity by postulating the existence of populations of “human-vocalization-neurons”, which are driven by inputs from lateral belt areas and AMY. Second, VA is not the only category-specific area that is selectively modulated by emotional valence. The fusiform face area shares the same feature. Emotional expressions were shown to modulate neural activity within the fusiform face area (intracranial recordings: Pourtois et al., 2009); fMRI: (Jehna et al., 2011; Harry et al., 2013), albeit less than within the face area in the superior temporal sulcus (Zhang et al., 2016). As revealed by a meta-analysis of fMRI studies, the fusiform face area appears to process emotional content for faces but not for other categories (i.e., visual scenes, Sabatinelli et al., 2011). Thus, predominantly auditory and predominantly visual cortices each comprise an area where category-specific processing – voices and faces, respectively – is modulated by emotional valence. In both cases, this emotion-linked encoding concerns stimuli of high social relevance. Third, the emotion-vocalization specific module within VA is very likely a stepping-stone towards a more global, hetero-modal representation of emotionally relevant information about people (Watson et al., 2014). The combined encoding of voices and faces, including emotional aspects, was shown to involve a small part of the superior temporal sulcus at the intersection of VA and the more posterior lying face area (Kreifelts et al., 2009; Ethofer et al., 2013).

Several aspects of the processing of emotional vocalizations, which remain to be explored, could be addressed in future studies with an event-related paradigm at 7T. This would allow to correlate the perceived valence by a given subject with the activation within a ROI and compare thus more precisely emotional modulation of vocalizations and non-vocalizations. This same design would be particularly adapted to investigate neural coupling between ROIs.

## Conclusions

Our results highlighted different stages in the processing of emotional vocalizations. Within the supratemporal plane, several lateral early-stage auditory areas responded strongly to non-verbal vocalizations and/or were modulated by emotional valence. However, none of these areas appeared to be dedicated to emotional processing of vocalizations only. This role was assumed by VA, where emotional valence modulated selectively responses to human vocalizations but not to other environmental sounds. In contrast, emotional valence modulated neural responses to both types of stimuli in right AMY. Correlation analysis revealed coupling between VA and early-stage auditory areas during the presentation of any vocalization, and between VA and right AMY during positive vocalizations. Thus, emotional vocalizations are processed in a dual pathway, whereby the emotion-vocalization module within VA integrates the input from the lateral early-stage auditory areas and from AMY.

**Funding:** This work was supported by the Swiss National Science Foundation Grant to S. Clarke (FNS 320030-159708) and by the Centre d'Imagerie BioMédicale (CIBM) of the UNIL, UNIGE, HUG, CHUV, EPFL and the Leenaards and Jeantet Foundations.

**Conflict of interest:** The authors have no conflict of interest.

**Ethical approval:** All procedures performed were in accordance and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards and ethical approval was obtained from the Ethical Committee of the Canton de Vaud (reference number 282/08).

## References

Aeschlimann M, Knebel J-F, Murray MM, Clarke S (2008) Emotional Pre-eminence of Human Vocalizations. Brain Topogr 20:239–248.

Amunts K, Kedo O, Kindler M, Pieperhoff P, Mohlberg H, Shah NJ, Habel U, Schneider F, Zilles K (2005) Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. Anat Embryol (Berl) 210:343–352.

Anderson AK, Christoff K, Stappen I, Panitz D, Ghahremani DG, Glover G, Gabrieli JDE, Sobel N (2003) Dissociated neural representations of intensity and valence in human olfaction. Nat Neurosci 6:196–202.

Arnal LH, Flinker A, Kleinschmidt A, Giraud A-L, Poeppel D (2015) Human Screams Occupy a Privileged Niche in the Communication Soundscape. Curr Biol 25:2051–2056.

Aubé W, Angulo-Perkins A, Peretz I, Concha L, Armony JL (2015) Fear across the senses: brain responses to music, vocalizations and facial expressions. Soc Cogn Affect Neurosci 10:399–407.

Baas D, Aleman A, Kahn RS (2004) Lateralization of amygdala activation: a systematic review of functional neuroimaging studies. Brain Res Rev 45:96–103.

Ball T, Derix J, Wentlandt J, Wieckhorst B, Speck O, Schulze-Bonhage A, Mutschler I (2009) Anatomical specificity of functional amygdala imaging of responses to stimuli with positive and negative emotional valence. J Neurosci Methods 180:57–70.

Ball T, Rahm B, Eickhoff SB, Schulze-Bonhage A, Speck O, Mutschler I (2007) Response Properties of Human Amygdala Subregions: Evidence Based on Functional MRI Combined with Probabilistic Anatomical Maps. PLOS ONE 2:e307.

Barbas H (2007) Flow of information for emotions through temporal and orbitofrontal pathways. J Anat 211:237–249.

Barbas H, Zikopoulos B, Timbie C (2011) Sensory Pathways and Emotional Context for Action in Primate Prefrontal Cortex. Biol Psychiatry 69:1133–1139.

Beaucousin V, Lacheret A, Turbelin M-R, Morel M, Mazoyer B, Tzourio-Mazoyer N (2007a) FMRI Study of Emotional Speech Comprehension. Cereb Cortex 17:339–352.

Beaucousin V, Lacheret A, Turbelin M-R, Morel M, Mazoyer B, Tzourio-Mazoyer N (2007b) FMRI study of emotional speech comprehension. Cereb Cortex N Y N 1991 17:339–352.

Belin P, Zatorre RJ, Ahad P (2002) Human temporal-lobe response to vocal sounds. Cogn Brain Res 13:17–26.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403:309–312.

Bestelmeyer PEG, Kotz SA, Belin P (2017) Effects of emotional valence and arousal on the voice perception network. Soc Cogn Affect Neurosci Available at: https://academic.oup.com/scan/article/doi/10.1093/scan/nsx059/3753554/Effects-of-emotional-valence-and-arousal-on-the [Accessed May 30, 2017].

Boubela RN, Kalcher K, Huf W, Seidel E-M, Derntl B, Pezawas L, Našel C, Moser E (2015) fMRI measurements of amygdala activation are confounded by stimulus correlated signal fluctuation in nearby veins draining distant brain regions. Sci Rep 5:10499.

Cammoun L, Thiran JP, Griffa A, Meuli R, Hagmann P, Clarke S (2015) Intrahemispheric cortico-cortical connections of the human auditory cortex. Brain Struct Funct 220:3537–3553.

Chevillet M, Riesenhuber M, Rauschecker JP (2011) Functional Correlates of the Anterolateral Processing Hierarchy in Human Auditory Cortex. J Neurosci 31:9345–9352.

Chiry O, Tardif E, Magistretti PJ, Clarke S (2003) Patterns of calcium-binding proteins support parallel and hierarchical organization of human auditory areas. Eur J Neurosci 17:397–410.

Clarke S, Morosan P (2012) Architecture, connectivity and transmitter receptors of human auditory cortex. In: In Poeppel, D., Overath, T., Popper, A. N., and Fay, R.R. (eds). (2012). Human Auditory Cortex., Springer Science+Business Media, LLC, New York.

Clarke S, Rivier F (1998) Compartments within human primary auditory cortex: evidence from cytochrome oxidase and acetylcholinesterase staining. Eur J Neurosci 10:741–745.

Costa VD, Lang PJ, Sabatinelli D, Versace F, Bradley MM (2010) Emotional Imagery: Assessing Pleasure and Arousal in the Brain’s Reward Circuitry. Hum Brain Mapp 31:1446–1457.

Costafreda SG, Brammer MJ, David AS, Fu CHY (2008) Predictors of amygdala activation during the processing of emotional stimuli: A meta-analysis of 385 PET and fMRI studies. Brain Res Rev 58:57–70.

Da Costa S, Zwaag W van der, Marques JP, Frackowiak RSJ, Clarke S, Saenz M (2011) Human Primary Auditory Cortex Follows the Shape of Heschl’s Gyrus. J Neurosci 31:14067–14075.

Da Costa S, Zwaag W van der, Miller LM, Clarke S, Saenz M (2013) Tuning In to Sound: Frequency-Selective Attentional Filter in Human Primary Auditory Cortex. J Neurosci 33:1858–1863.

Da Costa S, Bourquin NM-P, Knebel J-F, Saenz M, Zwaag W van der, Clarke S (2015) Representation of Sound Objects within Early-Stage Auditory Areas: A Repetition Effect Study Using 7T fMRI. PLOS ONE 10:e0124072.

Da Costa S, Clarke S, Crottaz-Herbette S (2018) Keeping track of sound objects in space: The contribution of early-stage auditory areas. Hear Res Available at: http://www.sciencedirect.com/science/article/pii/S0378595517306032.

De Meo R, Bourquin NM-P, Knebel J-F, Murray MM, Clarke S (2015) From bird to sparrow: Learning-induced modulations in fine-grained semantic discrimination. NeuroImage 118:163–173.

Di Virgilio G, Clarke S (1997) Direct interhemispheric visual input to human speech areas. Hum Brain Mapp 5:347–354.

Dietrich S, Hertrich I, Alter K, Ischebeck A, Ackermann H (2007) Semiotic aspects of human nonverbal vocalizations: a functional imaging study. Neuroreport 18:1891–1894.

Dietrich S, Hertrich I, Alter K, Ischebeck A, Ackermann H (2008) Understanding the emotional expression of verbal interjections: a functional MRI study. Neuroreport 19:1751–1755.

Ethofer T, Anders S, Erb M, Droll C, Royen L, Saur R, Reiterer S, Grodd W, Wildgruber D (2006a) Impact of voice on emotional judgment of faces: An event-related fMRI study. Hum Brain Mapp 27:707–714.

Ethofer T, Anders S, Wiethoff S, Erb M, Herbert C, Saur R, Grodd W, Wildgruber D (2006b) Effects of prosodic emotional intensity on activation of associative auditory cortex: NeuroReport 17:249–253.

Ethofer T, Bretscher J, Gschwind M, Kreifelts B, Wildgruber D, Vuilleumier P (2012) Emotional Voice Areas: Anatomic Location, Functional Properties, and Structural Connections Revealed by Combined fMRI/DTI. Cereb Cortex 22:191–200.

Ethofer T, Bretscher J, Wiethoff S, Bisch J, Schlipf S, Wildgruber D, Kreifelts B (2013) Functional responses and structural connections of cortical areas for processing faces and voices in the superior temporal sulcus. NeuroImage 76:45–56.

Ethofer T, Kreifelts B, Wiethoff S, Wolf J, Grodd W, Vuilleumier P, Wildgruber D (2008) Differential Influences of Emotion, Task, and Novelty on Brain Regions Underlying the Processing of Speech Melody. J Cogn Neurosci 21:1255–1268.

Ethofer T, Kreifelts B, Wiethoff S, Wolf J, Grodd W, Vuilleumier P, Wildgruber D (2009a) Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. J Cogn Neurosci 21:1255–1268.

Ethofer T, Van De Ville D, Scherer K, Vuilleumier P (2009b) Decoding of Emotional Information in Voice-Sensitive Cortices. Curr Biol 19:1028–1033.

Fecteau S, Belin P, Joanette Y, Armony JL (2007) Amygdala responses to nonlinguistic emotional vocalizations. NeuroImage 36:480–487.

Formisano E, Kim D-S, Salle FD, Moortele P-F van de, Ugurbil K, Goebel R (2003) Mirror-Symmetric Tonotopic Maps in Human Primary Auditory Cortex. Neuron 40:859–869.

Frühholz S, Grandjean D (2013) Processing of emotional vocalizations in bilateral inferior frontal cortex. Neurosci Biobehav Rev 37:2847–2855.

Frühholz S, Trost W, Grandjean D (2014) The role of the medial temporal limbic system in processing emotions in voice and music. Prog Neurobiol 123:1–17.

Frühholz S, Trost W, Kotz SA (2016) The sound of emotions—Towards a unifying neural network perspective of affective sound processing. Neurosci Biobehav Rev 68:96–110.

Ghashghaei HT, Barbas H (2002) Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. Neuroscience 115:1261–1279.

Goedhart AD, Van Der Sluis S, Houtveen JH, Willemsen G, De Geus EJC (2007) Comparison of time and frequency domain measures of RSA in ambulatory recordings. Psychophysiology 44:203–215.

Grandjean D, Sander D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P (2005) The voices of wrath: brain responses to angry prosody in meaningless speech. Nat Neurosci 8:145–146.

Hackett TA, Preuss TM, Kaas JH (2001) Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. J Comp Neurol 441:197–222.

Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield AQ (2002) Spectral and Temporal Processing in Human Auditory Cortex. Cereb Cortex 12:140–149.

Harry BB, Williams M, Davis C, Kim J (2013) Emotional expressions evoke a differential response in the fusiform face area. Front Hum Neurosci 7 Available at: https://www.frontiersin.org/articles/10.3389/fnhum.2013.00692/full [Accessed May 17, 2018].

Hurlemann R, Rehme AK, Diessel M, Kukolja J, Maier W, Walter H, Cohen MX (2008) Segregating intra-amygdalar responses to dynamic facial emotion with cytoarchitectonic maximum probability maps. J Neurosci Methods 172:13–20.

Janak PH, Tye KM (2015) From circuits to behaviour in the amygdala. Nature 517:284–292.

Jehna M, Langkammer C, Wallner-Blazek M, Neuper C, Loitfelder M, Ropele S, Fuchs S, Khalil M, Pluta-Fuerst A, Fazekas F, Enzinger C (2011) Cognitively preserved MS patients demonstrate functional differences in processing neutral and emotional faces. Brain Imaging Behav 5:241–251.

Kamali A, Sair HI, Blitz AM, Riascos RF, Mirbagheri S, Keser Z, Hasan KM (2016) Revealing the ventral amygdalofugal pathway of the human limbic system using high spatial resolution diffusion tensor tractography. Brain Struct Funct 221:3561–3569.

Kasper L, Bollmann S, Diaconescu AO, Hutton C, Heinzle J, Iglesias S, Hauser TU, Sebold M, Manjaly Z-M, Pruessmann KP, Stephan KE (2017) The PhysIO Toolbox for Modeling Physiological Noise in fMRI Data. J Neurosci Methods 276:56–72.

Keifer OP, Gutman DA, Hecht EE, Keilholz SD, Ressler KJ (2015) A comparative analysis of mouse and human medial geniculate nucleus connectivity: a DTI and anterograde tracing study. NeuroImage 105:53–66.

Kim S-G, Knösche TR (2016) Intracortical myelination in musicians with absolute pitch: Quantitative morphometry using 7-T MRI. Hum Brain Mapp 37:3486–3501.

Koelsch S (2010) Towards a neural basis of music-evoked emotions. Trends Cogn Sci 14:131–137.

Kreifelts B, Ethofer T, Shiozawa T, Grodd W, Wildgruber D (2009) Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus. Neuropsychologia 47:3059–3066.

Labar KS, Gitelman DR, Mesulam MM, Parrish TB (2001) Impact of signal-to-noise on functional Mri of the human amygdala. Neuroreport 12:3461–3464.

Latinus M, McAleer P, Bestelmeyer PEG, Belin P (2013) Norm-Based Coding of Voice Identity in Human Auditory Cortex. Curr Biol 23:1075–1080.

Lavan N, Rankin G, Lorking N, Scott S, McGettigan C (2017) Neural correlates of the affective properties of spontaneous and volitional laughter types. Neuropsychologia 95:30–39.

LeDoux JE, Ruggiero DA, Reis DJ (1985) Projections to the subcortical forebrain from anatomically defined regions of the medial geniculate body in the rat. J Comp Neurol 242:182–213.

Leitman DI, Wolf DH, Ragland JD, Laukka P, Loughead J, Valdez JN, Javitt DC, Turetsky BI, Gur RC (2010) “It’s Not What You Say, But How You Say it”: A Reciprocal Temporo-frontal Network for Affective Prosody. Front Hum Neurosci 4 Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2831710/ [Accessed March 21, 2016].

Liebenthal E, Silbersweig DA, Stern E (2016) The Language, Tone and Prosody of Emotions: Neural Substrates and Dynamics of Spoken-Word Emotion Perception. Front Neurosci 10 Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5099784/ [Accessed March 8, 2017].

Marques JP, Kober T, Krueger G, van der Zwaag W, Van de Moortele P-F, Gruetter R (2010) MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. NeuroImage 49:1271–1281.

Mathiak KA, Zvyagintsev M, Ackermann H, Mathiak K (2012) Lateralization of amygdala activation in fMRI may depend on phase-encoding polarity. Magn Reson Mater Phys Biol Med 25:177–182.

Meyer M, Zysset S, von Cramon DY, Alter K (2005) Distinct fMRI responses to laughter, speech, and sounds along the human peri-sylvian cortex. Brain Res Cogn Brain Res 24:291–306.

Moerel M, De Martino F, Formisano E (2014) An anatomical and functional topography of human auditory cortical areas. Front Neurosci 8 Available at: http://journal.frontiersin.org/article/10.3389/fnins.2014.00225/abstract [Accessed February 20, 2017].

Morris JS, Friston KJ, Büchel C, Frith CD, Young AW, Calder AJ, Dolan RJ (1998) A neuromodulatory role for the human amygdala in processing emotional facial expressions. Brain 121:47–57.

Morris JS, Scott SK, Dolan RJ (1999) Saying it with feeling: neural responses to emotional vocalizations. Neuropsychologia 37:1155–1163.

Murray RJ, Brosch T, Sander D (2014) The functional profile of the human amygdala in affective processing: Insights from intracranial recordings. Cortex 60:10–33.

Obleser J, Eisner F, Kotz SA (2008) Bilateral Speech Comprehension Reflects Differential Sensitivity to Spectral and Temporal Features. J Neurosci 28:8116–8123.

Obleser J, Zimmermann J, Van Meter J, Rauschecker JP (2007) Multiple Stages of Auditory Speech Perception Reflected in Event-Related fMRI. Cereb Cortex 17:2251–2257.

O'Doherty J, Rolls ET, Francis S, Bowtell R, McGlone F (2001) Representation of Pleasant and Aversive Taste in the Human Brain. J Neurophysiol 85:1315–1321.

Oldfield RC (1971) The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9:97–113.

Ottersen OP, Ben-Ari Y (1979) Afferent connections to the amygdaloid complex of the rat and cat. I. Projections from the thalamus. J Comp Neurol 187:401–424.

Pannese A, Grandjean D, Frühholz S (2016) Amygdala and auditory cortex exhibit distinct sensitivity to relevant acoustic features of auditory emotions. Cortex 85:116–125.

Pernet CR, McAleer P, Latinus M, Gorgolewski KJ, Charest I, Bestelmeyer PEG, Watson RH, Fleming D, Crabbe F, Valdes-Sosa M, Belin P (2015) The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. NeuroImage 119:164–174.

Phillips ML, Medford N, Young AW, Williams L, Williams SC, Bullmore ET, Gray JA, Brammer MJ (2001) Time courses of left and right amygdalar responses to fearful facial expressions. Hum Brain Mapp 12:193–202.

Phillips ML, Young AW, Scott SK, Calder AJ, Andrew C, Giampietro V, Williams SCR, Bullmore ET, Brammer M, Gray JA (1998) Neural responses to facial and vocal expressions of fear and disgust. Proc R Soc Lond B Biol Sci 265:1809–1817.

Plaisant O, Courtois R, Réveillère C, Mendelsohn GA, John OP (2010) Validation par analyse factorielle du Big Five Inventory français (BFI-Fr). Analyse convergente avec le NEO-PI-R. Ann Méd-Psychol Rev Psychiatr 168:97–106.

Pourtois G, Spinelli L, Seeck M, Vuilleumier P (2009) Modulation of Face Processing by Emotional Expression and Gaze Direction during Intracranial Recordings in Right Fusiform Cortex. J Cogn Neurosci 22:2086–2107.

Price JL, Amaral DG (1981) An autoradiographic study of the projections of the central nucleus of the monkey amygdala. J Neurosci Off J Soc Neurosci 1:1242–1259.

Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund H-J, Zilles K (2001) Probabilistic Mapping and Volume Measurement of Human Primary Auditory Cortex. NeuroImage 13:669–683.

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12:718–724.

Reddy RK, Ramachandra V, Kumar N, Singh NC (2009) Categorization of environmental sounds. Biol Cybern 100:299.

Reser DH, Burman KJ, Richardson KE, Spitzer MW, Rosa MGP (2009) Connections of the marmoset rostrotemporal auditory area: express pathways for analysis of affective content in hearing. Eur J Neurosci 30:578–592.

Rivier F, Clarke S (1997) Cytochrome Oxidase, Acetylcholinesterase, and NADPH-Diaphorase Staining in Human Supratemporal and Insular Cortex: Evidence for Multiple Auditory Areas. NeuroImage 6:288–304.

Roy AK, Shehzad Z, Margulies DS, Kelly AMC, Uddin LQ, Gotimer K, Biswal BB, Castellanos FX, Milham MP (2009) Functional connectivity of the human amygdala using resting state fMRI. NeuroImage 45:614–626.

Russchen FT (1982) Amygdalopetal projections in the cat. II. Subcortical afferent connections. A study with retrograde tracing techniques. J Comp Neurol 207:157–176.

Russchen FT, Amaral DG, Price JL (1987) The afferent input to the magnocellular division of the mediodorsal thalamic nucleus in the monkey, Macaca fascicularis. J Comp Neurol 256:175–210.

Sabatinelli D, Fortune EE, Li Q, Siddiqui A, Krafft C, Oliver WT, Beck S, Jeffries J (2011) Emotional perception: Meta-analyses of face and natural scene processing. NeuroImage 54:2524–2533.

Sander K, Brechmann A, Scheich H (2003) Audition of laughing and crying leads to right amygdala activation in a low-noise fMRI setting. Brain Res Protoc 11:81–91.

Sander K, Frome Y, Scheich H (2007) FMRI activations of amygdala, cingulate cortex, and auditory cortex by infant laughing and crying. Hum Brain Mapp 28:1007–1022.

Sander K, Scheich H (2005) Left Auditory Cortex and Amygdala, but Right Insula Dominance for Human Laughing and Crying. J Cogn Neurosci 17:1519–1531.

Sauter DA, Eisner F, Ekman P, Scott SK (2010) Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. Proc Natl Acad Sci 107:2408–2412.

Schirmer A, Gunter TC (2017) Temporal signatures of processing voiceness and emotion in sound. Soc Cogn Affect Neurosci 12:902–909.

Schirmer A, Kotz SA (2006) Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. Trends Cogn Sci 10:24–30.

Scott BH, Leccese PA, Saleem KS, Kikuchi Y, Mullarkey MP, Fukushima M, Mishkin M, Saunders RC (2015) Intrinsic Connections of the Core Auditory Cortical Regions and Rostral Supratemporal Plane in the Macaque Monkey. Cereb Cortex:bhv277.

Sergerie K, Chochol C, Armony JL (2008) The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies. Neurosci Biobehav Rev 32:811–830.

Shinonaga Y, Takada M, Mizuno N (1994) Direct projections from the non-laminated divisions of the medial geniculate nucleus to the temporal polar cortex and amygdala in the cat. J Comp Neurol 340:405–426.

Solano-Castiella E, Schäfer A, Reimer E, Türke E, Pröger T, Lohmann G, Trampel R, Turner R (2011) Parcellation of human amygdala in vivo using ultra high field structural MRI. NeuroImage 58:741–748.

Szameitat DP, Kreifelts B, Alter K, Szameitat AJ, Sterr A, Grodd W, Wildgruber D (2010) It is not always tickling: distinct cerebral responses during perception of different laughter types. NeuroImage 53:1264–1271.

Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, Dale AM (2004) Tonotopic Organization in Human Auditory Cortex Revealed by Progressions of Frequency Sensitivity. J Neurophysiol 91:1282–1296.

Tardif E, Clarke S (2001) Intrinsic connectivity of human auditory areas: a tracing study with DiI. Eur J Neurosci 13:1045–1050.

Tardif E, Probst A, Clarke S (2007) Laminar Specificity of Intrinsic Connections in Broca’s Area. Cereb Cortex 17:2949–2960.

van der Zwaag W, Gentile G, Gruetter R, Spierer L, Clarke S (2011) Where sound position influences sound object representations: A 7-T fMRI study. NeuroImage 54:1803–1811.

Viceic D, Campos R, Fornari E, Spierer L, Meuli R, Clarke S, Thiran J-P (2009) Local landmark-based registration for fMRI group studies of nonprimary auditory cortex. NeuroImage 44:145–153.

Viceic D, Fornari E, Thiran J-P, Maeder PP, Meuli R, Adriani M, Clarke S (2006) Human auditory belt areas specialized in sound recognition: a functional magnetic resonance imaging study. Neuroreport 17:1659–1662.

Viinikainen M, Kätsyri J, Sams M (2012) Representation of perceived sound valence in the human brain. Hum Brain Mapp 33:2295–2305.

Wallace MN, Johnston PW, Palmer AR (2002) Histochemical identification of cortical areas in the auditory region of the human brain. Exp Brain Res 143:499–508.

Warren JD, Scott SK, Price CJ, Griffiths TD (2006) Human brain mechanisms for the early analysis of voices. NeuroImage 31:1389–1397.

Watson R, Latinus M, Charest I, Crabbe F, Belin P (2014) People-selectivity, audiovisual integration and heteromodality in the superior temporal sulcus. Cortex 50:125–136.

Wiethoff S a b, Wildgruber D a, Grodd W b, Ethofer T a b (2009) Response and habituation of the amygdala during processing of emotional prosody. [Miscellaneous Article]. Neuroreport 20:1356–1360.

Wildgruber D, Ackermann H, Kreifelts B, Ethofer T (2006) Cerebral processing of linguistic and emotional prosody: fMRI studies. In: Progress in Brain Research (Anders S, Ende G, Junghofer M, Kissler J, Wildgruber D, eds), pp 249–268 Understanding Emotions. Elsevier. Available at: http://www.sciencedirect.com/science/article/pii/S0079612306560133 [Accessed March 20, 2019].

Wildgruber D, Ethofer T, Grandjean D, Kreifelts B (2009) A cerebral network model of speech prosody comprehension: International Journal of Speech-Language Pathology: Vol 11, No 4. Int J Speech Lang Pathol 11:277–281.

Wildgruber D, Hertrich I, Riecker A, Erb M, Anders S, Grodd W, Ackermann H (2004a) Distinct Frontal Regions Subserve Evaluation of Linguistic and Emotional Aspects of Speech Intonation. Cereb Cortex 14:1384–1389.

Wildgruber D, Hertrich I, Riecker A, Erb M, Anders S, Grodd W, Ackermann H (2004b) Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. Cereb Cortex N Y N 1991 14:1384–1389.

Winston JS, Gottfried JA, Kilner JM, Dolan RJ (2005) Integrated Neural Representations of Odor Intensity and Affective Valence in Human Amygdala. J Neurosci 25:8903–8907.

Wright, C.I., Fischer, H., Whalen, P.J., Mclnerney, S.C., Shin, L.M., Rauch, S.L. (2001) Wright_2001.pdf. NeuroReport 12:379–383.

Yukie M (2002) Connections between the amygdala and auditory cortical areas in the macaque monkey. Neurosci Res 42:219–229.

Zäske R, Awwad Shiekh Hasan B, Belin P (2017) It doesn't matter what you say: FMRI correlates of voice learning and recognition independent of speech content. Cortex 94:100–112.

Zhang H, Japee S, Nolan R, Chu C, Liu N, Ungerleider LG (2016) Face-selective regions differ in their ability to classify facial expressions. NeuroImage 130:77–90.

Zigmond, A. S., Snaith, R. P. (1983) The hospital anxiety and depression scale. Acta Psychiatr Scand 67:361–370.

Zilles K, Armstrong E, Schleicher A, Kretschmann H-J (1988) The human pattern of gyrification in the cerebral cortex. Anat Embryol (Berl) 179:173–179.

# STUDY B: Emotional value of the auditory space

Tiffany Grisendi, Stephanie Clarke and Sandra Da Costa

In preparation

# Emotional value of the auditory space

**Tiffany Grisendi[1], Stephanie Clarke[1] and Sandra Da Costa[2]**

[1] Service de Neuropsychologie et de Neuroréhabilitation, Centre Hospitalier Universitaire Vaudois (CHUV) and University of Lausanne, 1011 Lausanne, Switzerland

[2] Centre d'Imagerie Biomédicale, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

**Corresponding author:**
Dr. Sandra Da Costa
EPFL SB CIBM
CH F0 562 – Station 6
CH-1015 Lausanne
Tel.: +41 21 693 76 87
ORCID 0000-0002-8641-0494
Email: sandra.dacosta@epfl.ch

Number of pages: 28
Number of figures: 4
Number of tables: 3
Number of words in Abstract: 252

**Acknowledgements:** This work was supported by the Swiss National Science Foundation Grant to S. Clarke (FNS 320030-159708) and by the Centre d'Imagerie BioMédicale (CIBM) of the UNIL, UNIGE, HUG, CHUV, EPFL and the Leenaards and Jeantet Foundations.

**Conflict of interest:** The authors declare no competing financial interests.

**Sharing data:** The dataset of this manuscript is available upon request to the corresponding author.

## Abstract

Evidence from behavioural studies suggests that the spatial origin of sounds may influence the perception of emotional valence. Using 7T fMRI we have investigated the impact of the type of sound (vocalizations; non-vocalizations), emotional valence (positive, neutral, negative) and spatial origin (left, centre, right) on the encoding in early-stage auditory areas and in the voice area. The combination of these different characteristics resulted in a total of 18 conditions (2 Types x 3 Valences x 3 Lateralizations), which were presented in a pseudo-randomized order in blocks of eleven different sounds (of the same condition) in 12 distinct runs of 6min. In addition, the subjects (N = 14, with normal hearing) also listened to two different localizers (a tonotopy paradigm and a voice localizer), which were used to define the regions of interest. A 3-way repeated measure ANOVA on the BOLD responses revealed bilateral significant effects and interactions in the primary auditory cortex, the lateral early-stage auditory areas, and the voice area. Positive vocalizations presented on the left side yielded greater activity in bilateral primary auditory cortex than did neutral or negative vocalizations or any other stimuli at any of the three positions. The voice area did not share the same preference for the left space; spatial attributes modulated its activation by sound objects conveying positive or neutral emotional valence when presented on the right or left side (but not at the centre). Comparison with a previous study indicates that spatial cues may render emotional valence more salient within the early-stage auditory areas.

**Keywords:**

**Human vocalizations, emotions, auditory belt areas, voice area, lateralization, 7T fMRI**

## Abbreviations

| | |
|---|---|
| AI | primary auditory area |
| HVN | human vocalizations with negative emotional valence |
| HVP | human vocalizations with positive emotional valence |
| HV0 | human vocalizations with neutral emotional valence |
| NVN | non-vocalizations with negative emotional valence |
| NVP | non-vocalizations with positive emotional valence |
| NV0 | non-vocalizations with neutral emotional valence |
| R | rostral (primary) auditory area |
| VA | voice area |

## Introduction

Three lines of evidence suggest that the spatial origin of sounds influences the perception of emotional valence. First, looming sounds tend to be perceived as more unpleasant, potent, arousing and intense than receding sounds (Bach et al., 2008, 2009; Tajadura-Jiménez et al., 2010b). Second, sounds were reported to be more arousing when presented behind than in front of a person and this effects was stronger for natural sounds, such as human or animal vocalizations, than tones (Tajadura-Jiménez et al., 2010a). Third, when presented in a dichotic paradigm emotional vocalizations were shown to yield asymmetrical behavioural scores. An early study used syllables without significance spoken in seven different emotional intonations. The performance in detecting one emotion, defined as target, was significantly better for stimuli presented to the left than the right ear (Erhan et al., 1998). A later study used four words, which differed in the initial consonant, and which were spoken in four different emotional intonations. The subjects attended either both ears or one of them at a time. Performance analysis revealed a significant left-ear advantage for identifying the emotion (Jäncke et al., 2001). The behavioural results of either study were interpreted in terms of right hemispheric competence for emotional processing (for recent review e. g. (Gadea et al., 2011)). The alternative interpretation, that the emotional perception may be modulated by the lateralization of the sound, as it is for looming vs. receding sounds (Bach et al., 2008, 2009; Tajadura-Jiménez et al., 2010b), has not been considered.

The encoding of the auditory space is believed to be partially independent of the encoding of sound meaning. A series of seminal studies lead to the formulation of the dual-stream model of auditory processing, which posits partially independent encoding of sound meaning along the anterior temporal convexity and that of sound position on the parietal convexity ((Anourova et al., 2001; Maeder et al., 2001; Hart et al., 2004; Ahveninen et al., 2006; De Santis et al., 2007a, 2007b; Rauschecker and Scott, 2009); for a review (Arnott et al., 2004)). The functional independence of the two pathways has been documented in patient studies, where lesions limited to the ventral stream impaired sound recognition but not localization and conversely lesions limited to the dorsal stream impaired sound localization but not recognition (Clarke et al., 2000, 2002; Rey et al., 2007).

Recent evidence indicates that the combined encoding of sound object identity and location involves a separate, third processing stream, referred to also as the lateral pathway (Clarke and Geiser, 2015). Its initial demonstration relied on repetition priming paradigms; neural populations, which encoded the combined representation, displayed repetition enhancement when an object changed position and repetition suppression when it did not, both in EEG (Bourquin et al., 2013) and in 7T fMRI experiments (Da Costa et al., 2018). The latter identified several early-stage auditory areas on the supratemporal plane which participate in the combined encoding of sound object identity and position. The position-linked representation of sound objects, as supported by the lateral auditory pathway, is likely to contribute to auditory streaming, where spatial cues play an important role in the very early processing stages (Eramudugolla et al., 2008). The functional independence of the lateral and dorsal auditory pathways, has been demonstrated in patient studies, where the implicit use of auditory spatial cues was preserved for the segregation of sound objects, despite severe sound localization deficits, including cortical spatial deafness (Thiran and Clarke, 2003; Duffour-Nikolov et al., 2012; Tissieres et al., 2019).

The early-stage primary and non-primary auditory areas are located on the supratemporal plane and constitute first steps of cortical processing; several of them were defined by anatomical, histological and/or functional markers in post-mortem studies and by functional criteria (Clarke and Morosan, 2012). The primary auditory cortex is roughly co-extensive with Heschl’s gyrus (Zilles et al., 1988; Rademacher et al., 2001) and consists of two orderly tonotopic representations (Formisano et al., 2003; Da Costa et al., 2011, 2014; Moerel et al., 2014). The surrounding plana polare and temporale comprise several non-primary auditory areas, which were characterized on the basis of histological criteria (Rivier and Clarke, 1997; Clarke and Rivier, 1998; Hackett et al., 2001; Wallace et al., 2002; Chiry et al., 2003). Their Talairach coordinates were used in activation studies (Viceic et al., 2006; van der Zwaag et al., 2011; Besle et al., 2019), in addition to the identification of the primary auditory cortex by means of tonotopic mapping (Da Costa et al., 2011, 2015, 2018).

Human vocalizations constitute emotionally highly potent stimuli. They are processed in a dedicated region on the superior temporal gyrus, the voice area (VA), which is defined by its stronger response to human than animal vocalizations (Belin et al., 2000). The encoding of vocalizations within VA is modulated by emotional valence, as demonstrated in a series of seminal studies (Belin et al., 2002; Grandjean et al., 2005; Ethofer et al., 2006, 2008, 2009, 2012; Beaucousin et al., 2007; Obleser et al., 2007, 2008; Bestelmeyer et al., 2017). In addition to VA, the emotional valence of vocalizations impacts also the activity on Heschl’s gyrus and the antero-lateral part of the planum temporale (Wildgruber et al., 2005; Leitman et al., 2010; Ethofer et al., 2012; Arnal et al., 2015; Lavan et al., 2017) . The relatively low spatial resolution used in these studies did not allow to analyse separately neural activity within VA and within individual auditory areas. This has been done in a recent 7T fMRI study, which used human vocalizations and non-vocalizations with positive, neutral or negative valence (Grisendi et al., 2019). Several early-stage auditory areas yielded stronger responses to non-verbal vocalizations and/or were modulated by emotional valence. In contrast, in

VA emotional valence modulated selectively responses to human vocalizations but not to non-vocalizations.

The impact of emotional valence on the processing in respect to the ventral and dorsal auditory streams was investigated in a fMRI study (Kryklywy et al., 2013). Neural activity elicited by environmental sounds (predominantly human vocalizations) with positive, neutral or negative valence, presented at one of two left or two right positions, yielded a main effect of position bilaterally in a temporo-parietal region driven by a stronger activity to contralateral stimuli. A main effect of emotion was present bilaterally in an antero-superior temporal region, driven by stronger activity to emotional than neutral stimuli. The right auditory cortex yielded a significant interaction, driven by stronger response to contralateral positive stimuli. In a follow-up study (Kryklywy et al., 2018) the data were analysed with multi-voxel pattern analysis, which revealed overlapping representations of spatial and emotional attributes within the posterior part of the supratemporal plane.

In summary, human vocalizations convey strongly emotional valence, with a major involvement of VA and of the postero-lateral part of the planum temporale (Wildgruber et al., 2005; Leitman et al., 2010; Ethofer et al., 2012; Arnal et al., 2015; Lavan et al., 2017). The perceived emotional valence of sounds, including vocalizations, is modulated by spatial attributes as demonstrated for looming sounds (Bach et al., 2008, 2009; Tajadura-Jiménez et al., 2010b). A likely candidate for the interaction between emotional valence and spatial attributes of sounds is the planum temporale (Kryklywy et al., 2018). It is currently unclear whether other spatial attributes, such as left vs. right locations (and not simply left vs. right ear), modulate emotional perception and its encoding as well, and whether human vocalizations vs. other environmental sounds differ in this respect.

We have addressed these issues and hypothetized that specific early-stage auditory areas and/or VA may display one or several of the following characteristics:

i) The encoding of emotional vocalizations is more strongly modulated by their position than that of neutral vocalizations or non-vocalizations;
ii) The encoding of emotional valence is modulated by the spatial origin of the sound;
iii) The spatial origin of the sound impacts differently the encoding of vocalizations vs. non-vocalizations.

Furthermore, we expected to find spatial, emotional and vocalization selectivity, as reported in previous studies (Belin et al., 2002; Grandjean et al., 2005; Wildgruber et al., 2005; Ethofer et al., 2006, 2008, 2009, 2012; Beaucousin et al., 2007; Obleser et al., 2007, 2008; Leitman et al., 2010; Kryklywy et al., 2013; Arnal et al., 2015; Bestelmeyer et al., 2017; Lavan et al., 2017; Da Costa et al., 2018; Grisendi et al., 2019). To test these three hypotheses, we have made use of the high spatial resolution of ultra-high field fMRI at 7T to investigate the representation of human vocalizations *vs*. other environmental sounds, and their modulation by emotional valence and/or by their position within early-stage auditory areas and VA.

## Materials and Methods

### *Subjects*

Fourteen subjects (9 female, 12 right-handed, mean age 26.36 ± 4.19 years) participated in this study. Due to problem during the acquisition, one subject was discarded. All were French native speakers, without musical training. None reported history of neurological or psychiatric illness or hearing deficits and all had hearing thresholds within normal limits. Prior to the imaging session, each subject had to complete six questionnaires on their health status, handedness (Edinburgh Handedness Inventory, (Oldfield, 1971), anxiety and depression state (Hospital Anxiety and Depression, HAD, scale; (Zigmond and Snaith, 1983), personality traits (Big-Five Inventory, (Courtois et al., 2018), and a musical aptitude questionnaire developed in the lab. These questionnaires revealed no significant differences in personality traits nor in mood disorders. Thus, our group of volunteers was representative of the normal population and no subject was excluded of the study. The experimental procedures were approved by the Ethics Committee of the Canton de Vaud; all subjects gave written, informed consent.

### *Experimental design and statistical analysis*

The experimental design consisted of two fMRI sessions (~55-60 min each) during which auditory stimuli were presented while the subjects passively listen to the stimuli with eyes closed. In total, each subject performed two runs of tonotopy mappings, one run of voice localizer, and twelve runs of “emotions&space” runs. The latter consisted of 20s of silent rest (with no auditory stimuli except the scanner noise), followed by nine 36s-blocks of eleven sounds of the same condition (22s sounds and 14s of silent rest), and again 20s of silent rest. Each block was composed of eleven different sounds from the same category (human vocalizations or other environmental sounds), all of which had the same emotional valence (positive, neutral or negative) and the same lateralization (left, centre, right). Finally, blocks and their sequence order were pseudo-randomized within runs and across subjects.

Sounds (16 bits, stereo, sampling rate of 41kHz) presented binaurally at 80 ± 8 dB SPL via MRI-compatible headphones (SensiMetrics S14, SensiMetrics, USA), with a prior filtering with the SensiMetrics filters to obtain a flat frequency transmission, using MATLAB (R2015b, The

MathWorks, Inc., Natick, Massachusetts, United States) and the Psychophysics Toolbox (www.psychtoolbox.org). The auditory stimuli were the same as the battery used in previous studies (Aeschlimann et al., 2008; Grisendi et al., 2019), the total 66 different emotional sound files were 2s-long and were equally distributed in the six categories: Human Vocalizations Positive (HVP), Human Vocalizations Neutral (HV0), Human Vocalizations Negative (HVN), Non-Vocalizations Positive (NVP), Non-Vocalizations Neutral (NV0), and Non-Vocalizations Negative (NVN). Sounds were lateralized by creating artificially a temporal shift of 0.3s between the left and right channel (corresponding to ~60), using the available software Audacity (Audacity Team, https://audacityteam.org), and were either perceived as presented on the left, the centre or the right auditory space. Thus, the combination of all the different characteristics resulted in a total of 18 conditions (2 Types x 3 Valences x 3 Lateralizations).

As previously, using a specific software, PRAAT (http://www.fon.hum.uva.nl/praat/), and MATLAB scripts, the sound acoustic characteristics (spectrograms, mean fundamental frequency, mean intensity, harmonics to noise ratio, power, centre of gravity, mean Wiener entropy and spectral structure variation) were controlled for each category: first, the significant differences between the mean spectrogram of pairs of sounds of different categories were maintained < 1% to avoid bias towards a specific category (as in De Meo et al., 2015); second, all the sounds characteristics were tested with a 2-way repeated measures ANOVA with the factors Vocalization (Human-Vocalizations, Non-Vocalizations) x Valence (Positive, Neutral, Negative) to compare the effect of each acoustic feature on the sound categories. As already reported in our previous study (Grisendi et al., 2019), the analysis on mean Wiener entropy showed a main effect of Vocalizations [F(1,64) = 18.68, p = 0.0015], a main effect of Valence [F(2,63) = 21.14, p = 1.17E-5] and an interaction Vocalizations x Valence [F(2,63) = 8.28, p = 0.002]; while the same analysis on the centre of gravity revealed a main effect of Valence [F(2,63) = 10.51, p = 0.0007]. The analysis of the harmonics-to-noise ratios highlighted a main effect of Vocalizations [F(1,64) = 134.23, p = 4.06E-7], a main effect of Valence [F(2,63) = 69,61, p = 9.78E-10] and an interaction of Vocalizations x Valence [F(2,63) = 17.91, p = 3.48E-5], and these of the power showed an interaction of Vocalizations x Valence on the mean intensity [F(2,63) = 12.47, p = 0.0003] and on the power [F(2,63) = 14.77, p = 0.0001].

*Regions of interest definition*

The subdivision of the early-stage auditory areas was based on the individual frequency preferences as described previously (Da Costa et al., 2015, 2018), the subjects listen to two runs (one ascending and one descending) of a modified version of a tonotopic mapping paradigm (as in previous studies (Da Costa et al., 2011, 2013, 2015, 2018)), which consisted of progressions of 2s-bursts of pure tones (14 frequencies, between 88 and 8000 Hz, in half octave steps) presented in 12 identical cycles of 28s followed by a 12-s silent pause for a total duration of 8 min. Then, briefly, based on the resulting individual frequency reversals and anatomical landmarks, each early-stage auditory area was localized and defined in each subject as the primary auditory cortex, A1 and R, as well as the lateral (L1, L2, L3, L4) and medial non-primary areas (M1, M2, M3, M4). The coordinates of these regions were in accordance with previously published values (Table 1; (Viceic et al., 2006; van der Zwaag et al., 2011; Da Costa et al., 2015, 2018).

**Table 1: Mean MNI coordinates (center of gravity) of all ROIs. STD = standard deviation**

| ROI | X | ± | STD(X) | Y | ± | STD(Y) | Z | ± | STD(Z) |
|---|---|---|---|---|---|---|---|---|---|
| ***Left hemisphere*** | | | | | | | | | |
| **A1** | -45.20 | ± | 5.61 | -28.74 | ± | 5.42 | 9.11 | ± | 4.04 |
| **R** | -42.84 | ± | 4.84 | -21.79 | ± | 5.11 | 7.50 | ± | 4.37 |
| **L1** | -57.25 | ± | 5.74 | -37.57 | ± | 7.71 | 18.06 | ± | 8.46 |
| **L2** | -58.14 | ± | 5.55 | -21.17 | ± | 6.22 | 6.76 | ± | 5.49 |
| **L3** | -52.89 | ± | 5.41 | -9.08 | ± | 6.75 | 0.60 | ± | 4.67 |
| **L4** | -45.23 | ± | 4.38 | -4.18 | ± | 11.06 | -11.17 | ± | 7.63 |
| **M1** | -46.51 | ± | 5.74 | -38.74 | ± | 4.57 | 23.72 | ± | 7.95 |
| **M2** | -36.03 | ± | 2.71 | -33.44 | ± | 2.85 | 17.74 | ± | 3.36 |
| **M3** | -33.14 | ± | 2.80 | -29.26 | ± | 2.50 | 17.48 | ± | 3.25 |
| **M4** | -35.80 | ± | 3.17 | -14.80 | ± | 9.49 | -2.84 | ± | 11.85 |
| **VA** | -55.50 | ± | 6.47 | -33.46 | ± | 10.53 | 6.08 | ± | 5.62 |
| ***Right Hemisphere*** | | | | | | | | | |
| **A1** | 49.54 | ± | 5.19 | -23.74 | ± | 5.08 | 10.60 | ± | 3.49 |
| **R** | 45.49 | ± | 4.65 | -17.56 | ± | 4.88 | 6.73 | ± | 4.83 |
| **L1** | 60.92 | ± | 5.04 | -30.13 | ± | 4.67 | 21.69 | ± | 9.94 |
| **L2** | 62.40 | ± | 4.10 | -18.00 | ± | 7.27 | 7.07 | ± | 4.57 |
| **L3** | 55.99 | ± | 5.43 | -4.77 | ± | 6.81 | -0.24 | ± | 4.68 |
| **L4** | 46.90 | ± | 4.63 | -0.42 | ± | 10.06 | -11.67 | ± | 6.99 |
| **M1** | 48.99 | ± | 6.26 | -31.56 | ± | 3.64 | 26.70 | ± | 8.41 |
| **M2** | 38.04 | ± | 3.49 | -29.90 | ± | 3.19 | 18.33 | ± | 3.83 |
| **M3** | 35.14 | ± | 3.12 | -26.33 | ± | 3.43 | 16.28 | ± | 4.06 |
| **M4** | 34.95 | ± | 2.92 | -10.74 | ± | 10.96 | -3.30 | ± | 10.32 |
| **VA** | 48.79 | ± | 7.60 | -31.39 | ± | 7.37 | 5.46 | ± | 4.98 |

Finally, the localization VA was defined using a specific voice localizer used in previous studies (Belin et al., 2002; Pernet et al., 2015). Briefly, human vocalizations (vowels, words, syllables laughs, sighs, cries, coughs, etc.) and environmental sounds (falls, wind, animals sounds, etc.) were presented in a 10-min run, which consisted of forty 20s-long blocks (with 8s of sounds followed by a silent pause of 12s). This localizer was developed to easily and consistently identify the individual voice area along the lateral side of temporal plane, by displaying the results of the general linear model (GLM) contrast Human vocalizations *vs*. Environmental sounds. In this study, the same approach was used in BrainVoyager (BrainVoyager 20.6 for Windows, Brain Innovation, Maastricht, Netherlands). After initial preprocessing, the functional run was first aligned with the subject anatomical, and analysed with a general linear model using a boxcar design for the two conditions. Second, the results of the contrast Human vocalization vs. Environmental sounds was projects on the individual 3D volume rendering with a p value of $p < 0.005$ (uncorrected) in order to cover the same extend in each subject. Finally, the activated region within the bilateral lateral borders of the STS/STG was manually selected as a patch of interest using the manual drawing tools from BrainVoyager and projected back into the MNI space and saved as the individual region of interest. The coordinates of the VA were also in accordance with those of previous studies (Table 1; Belin et al., 2002; Pernet et al., 2015).

*Imaging parameters and data analysis*

Brain imaging was acquired on a 7-Tesla MRI scanner (Siemens MAGNETOM scanner, Siemens Medical Solutions, Germany) with an 32-channel head RF-coil (Nova Medical Inc., MA, USA). Functional datasets were obtained with a 2D-EPI sinusoidal simultaneous multi-slice sequence (1.5 x 1.5 mm in-plane resolution, slice thickness = 1.5 mm, TR = 2000 ms, TE = 23 ms, flip angle = 90°, slice gap = 0 mm, matrix size = 146 x 146, field of view = 222 x 222, with 40 oblique slices covering the superior temporal plane). T1-weigthed 3D structural images were obtained with a MP2RAGE sequence (resolution = 0.6 x 0.6 x 0.6 $mm^3$, TR = 6000 ms, TE = 4.94 ms, TI1/TI2 = 800/2700 ms, flip angle 1/flip angle 2 = 7/5, slice gap = 0 mm, matrix size = 320 x 320, field of view = 192 x 192 (Marques et al., 2010)). Finally, the physiological noise (respiration and heart beat) was recorded during the experiment using a plethysmograph and respiratory belt provided from the MRI scanner vendor.

The data was processed with BrainVoyager with the following steps: scan time correction (except for tonotopic mappings runs), temporal filtering, motion correction, segmentation and normalization into the MNI space. Individual frequency preferences were extracted with a linear cross-correlation analysis, resulting correlation maps were averaged together (ascending and descending correlation map) to define the best frequency value for each voxel in the volumetric space, and then the average map was projected onto the cortical surface meshes for the ROIs definition (Da Costa et al., 2011, 2013, 2015, 2018). For the VA localizer and the emotion&space runs, a random effects (RFX) analysis was performed at the group level, with movement and respiration parameters as regressor, and then we tested for the contrast 'Sounds vs. Silence' with an FDR correction at $q < 0.05$ (p <

0.05). The  GLM results for the VA localizer was used to define bilateral individual Vas, while the GLM results for the emotion&space runs were used to verify that our ROIs were activated by the paradigm. The scope of this paper was to evaluate the effects of spatial origin on the encoding of emotional sounds, therefore the remaining analysis focused on the BOLD responses extracted from all the ROIs.

Functional individual BOLD time courses were processed as the following: first, they were extracted using BrainVoyager, imported into MATLAB. Second, they were normalized by their own mean signal, and divided according to their condition. Third, they were averaged spatially (across all voxels within each ROI), temporally (over blocks and runs), and across the 13 subjects. The resulting time course consisted of 18 time points for each ROI and condition. Finally, these time courses were analysed with a time-point-by-time-point 3-Way repeated measure ANOVA, 2 Vocalization (Human-Vocalizations, Non-Vocalizations) x 3 Valence (Positive, Neutral, Negative) x 3 Lateralization (Left, Centre, Right) according to (Da Costa et al., 2015, 2018; Grisendi et al., 2019). This 3-way ANOVA was further decomposed for each vocalization type onto a 2-way repeated measure ANOVA, 3 Valence (Positive, Neutral, Negative) x 3 Lateralization (Left, Centre, Right). For each ANOVA, and each pair of condition, post hoc time-point-by-time-point paired t-tests were performed to evaluate the causality of the effects. Finally, results were restricted temporally by only considering at least three consecutive time points with significant p-values lower or equal to 0.05.

### *Physiological noise processing*

Heartbeat and respiration recordings were processed with an open-source toolbox for Matlab, TAPAS PhysIO (Kasper et al., 2017). The cardiac rates were further analysed with the same pipeline than the BOLD responses to obtain a pulse time course for each condition, while the respiration rates were used within the GLM model as motion regressor. As in Grisendi et al. (2019), the effect of space and emotional contents of the sounds on the individual cardiac rhythm was evaluated by computing the heart rate variability according to (Goedhart et al., 2007).

## Results

To explore how far emotional valence and/or position modulate the encoding of vocalizations vs. non-vocalizations within specific ROIs, we have analysed the BOLD responses within each area with a 3-way repeated measure ANOVA with factors Valence (Positive, Neutral, Negative), Lateralization (Left, Centre, Right) and Vocalization (Human-Vocalizations, Non-Vocalizations). The significance of main effects and interactions within individual early-stage auditory areas and within VA (Figs 1 and 2) provided answers for the three hypotheses we set out to test.

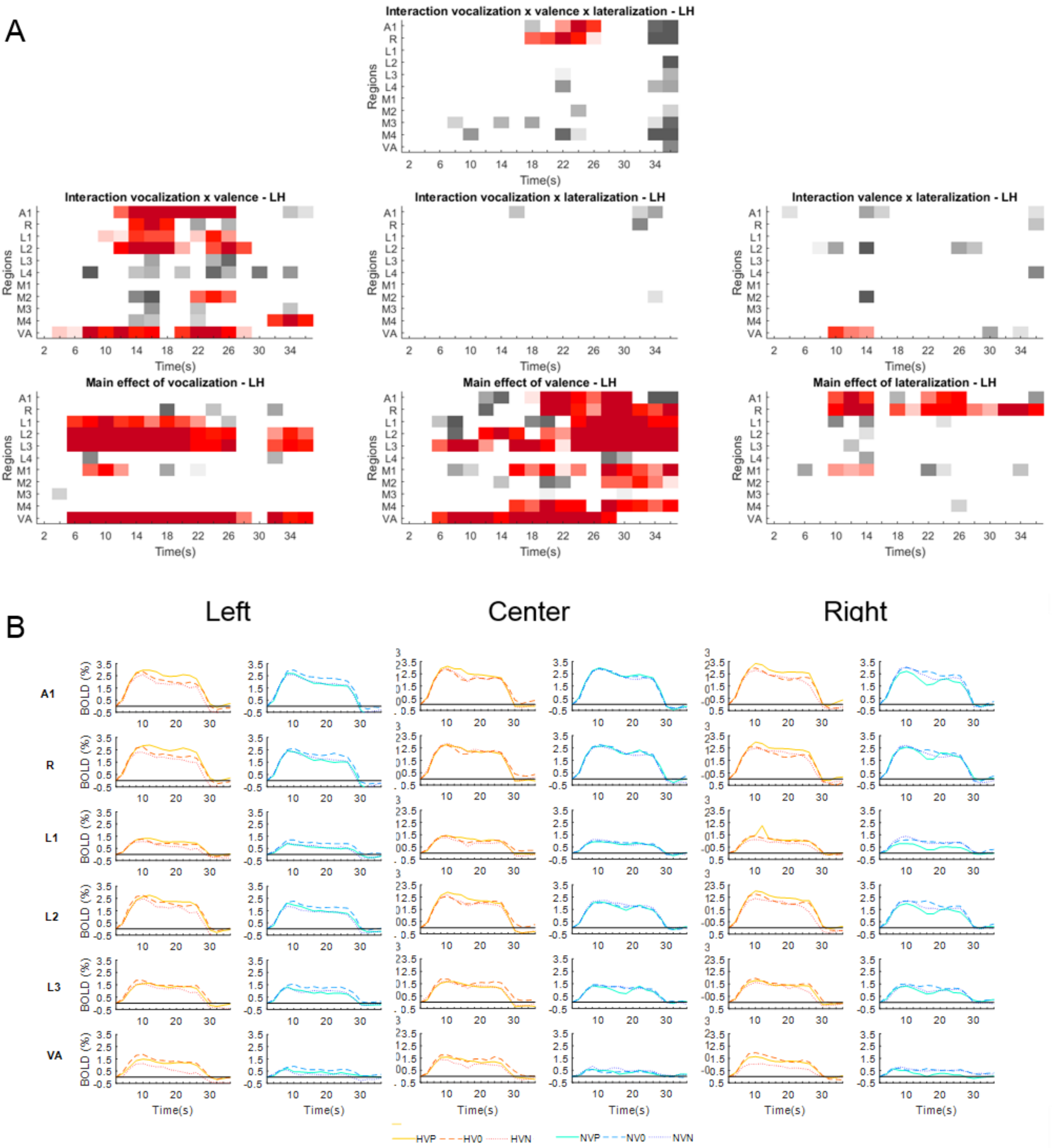

**Figure 1. Activations elicited in the left hemisphere**. A. Statistical analysis of the BOLD signal by means of a 2-way ANOVA with factors Vocalization (vocalizations, non-vocalizations) x Valence (positive, neutral, negative) x Lateralization (left, centre, right). The ROIs, .i. e., early-stage auditory areas and VA, are represented on the y-axis, the time points on the x-axis; red indicates a p-value lower or equal to 0.05 for at least three consecutive time points, grey a p-value lower or equal to 0.05 for isolated time-points. LH = left hemisphere. **B.** BOLD time courses for selected early-stage areas and VA, presented on the left, at the centre or on the right. Human vocalization categories are depicted in orange [HVP (solid line), HV0 (dashed line), HVN (dotted line)] non-vocalization categories in blue [NVP (solid line), NV0 (dashed line), NVN (dotted line)]. Full line denotes positive, interrupted line neutral and dotted line negative valence.

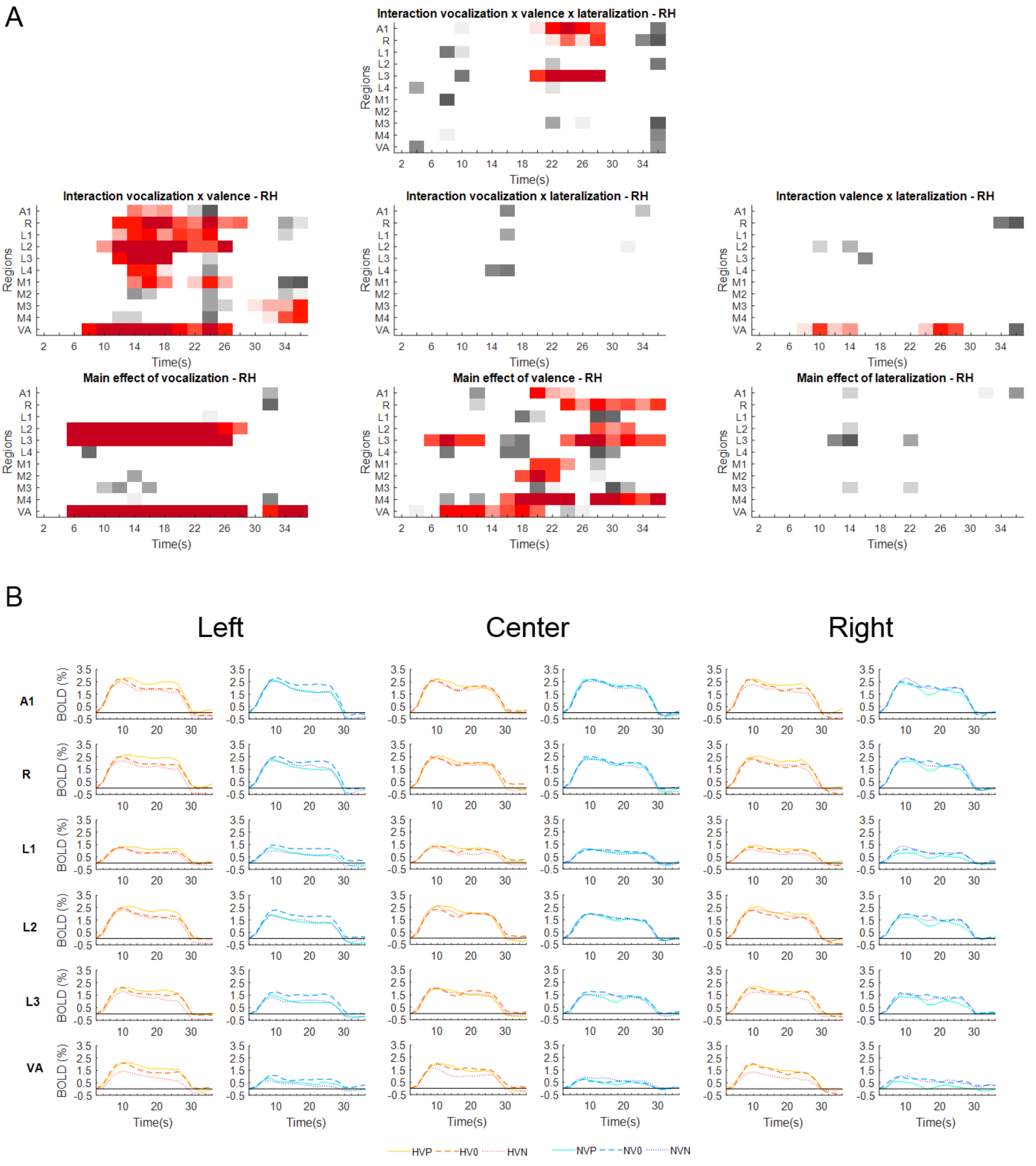

**Figure 2. Activations elicited in the right hemisphere. A.** Statistical analysis of the BOLD signal by means of a 2-way ANOVA with factors Vocalization (vocalizations, non-vocalizations) x Valence (positive, neutral, negative) x Lateralization (left, centre, right). The ROIs, .i. e., early-stage auditory areas and VA, are represented on the y-axis, the time points on the x-axis; red indicates a p-value lower or equal to 0.05 for at least three consecutive time points, grey a p-value lower or equal to 0.05 for isolated time-points. RH = right hemisphere. **B.** BOLD time courses for selected early-stage areas and VA, presented on the left, at the centre or on the right. Human vocalization categories are depicted in orange [HVP (solid line), HV0 (dashed line), HVN (dotted line)] non-vocalization categories in blue [NVP (solid line), NV0 (dashed line), NVN (dotted line)]. Full line denotes positive, interrupted line neutral and dotted line negative valence.

***The encoding of emotional vocalizations is more strongly modulated by their position than that of neutral vocalizations or non-vocalizations*** *(hypothesis i)*

The triple interaction Vocalization x Valence x Lateralization was significant in A1 and R in the left hemisphere and in A1, R and L3 in the right hemisphere. In *left A1* the significant time window was 22-26 s post-stimulus onset. During this time window the triple interaction was driven by two double interactions (Table 2). First, the interaction Vocalization x Valence was significant for stimuli presented on the left (but not right or at the centre). Second, the interaction Vocalizations x Lateralization was significant for positive (but not neutral or negative) stimuli. These interactions were driven by the significant main effect of Vocalization for positive stimuli presented on the left, vocalizations yielding stronger activation than non-vocalizations. Post-hoc comparisons revealed during the same time window that among the vocalizations presented on the left positive ones yielded significantly greater activation than neutral or negative ones. Thus, taken together these results highlight the pro-eminence of positive vocalizations when presented on the left.

In *left R* the significant time window was 18-26 s post-stimulus onset. During this time window the triple interaction was driven by two double interactions (Table 2). First, the interaction Vocalization x Valence was significant for stimuli presented on the left (but not right or at the centre). Second, the interaction Vocalizations x Lateralization was significant for positive (but not neutral or negative) stimuli. These two interactions were driven by the significant main effect of Vocalization for positive stimuli presented on the left, vocalizations yielding stronger activation than non-vocalizations. Post-hoc comparisons revealed during the same time window that among the vocalizations presented on the left positive ones yielded significantly greater activation than neutral or negative ones. Also positive vocalizations yielded significantly stronger activation when presented on the left than at the centre or on the right. Thus, taken together these results highlight the pro-eminence of positive vocalizations when presented on the left.

In *right A1* the significant time window was 20-28 s post-stimulus onset. During this time window the triple interaction was driven by two double interactions (Table 2). First, the interaction Vocalization x Valence was significant for stimuli presented on the left (but not right or at the centre). Second, the interaction Vocalizations x Lateralization was significant for positive (but not neutral or negative) stimuli. Post-hoc comparisons revealed during the same time window that among the vocalizations presented on the left positive ones yielded significantly greater activation than neutral or negative ones. Also positive vocalizations yielded significantly stronger activation when presented on the left than at the centre or on the right. Thus, taken together these results highlight the pro-eminence of positive vocalizations when presented on the left.

**Table 2. Summary of significant triple interaction Vocalization x Valence x Lateralization and the ensuing double interactions and main effects in individual ROIs of the left and right hemispheres. For the time window of significant triple interaction within a given ROI are listed the related double interactions and main effects.**

| **ROI with significant triple interaction Vocalization x Valence x Lateralization (time window of significance)** | **Significant related double interaction during the same time window** | **Significant related main effect during the same time window** |
|---|---|---|
| ***Left hemisphere*** | | |
| A1 (22 - 26 s) | Vocalization x Valence for left<br>Vocalization x Lateralization for positive | Vocalization for positive on left (vocalizations > non-vocalizations |
| R (18 - 26 s) | Vocalization x Valence for left<br>Vocalization x Lateralization for positive | Vocalization for positive on left (vocalizations > non-vocalizations |
| ***Right hemisphere*** | | |
| A1 (20 – 28 s) | Vocalization x Valence for left<br>Vocalization x Lateralization for positive | |
| R (20 – 28 s) | Vocalization x Valence for left<br>Vocalization x Lateralization for positive | Vocalization for positive on left (vocalizations > non-vocalizations |
| L3 (20 – 28 s) | Vocalization x Valence for left<br>Vocalization x Valence for right<br>Vocalization x Lateralization for positive<br>Valence x Lateralization for vocalizations<br>Valence x Lateralization for non-vocalizations | Vocalization for positive on left (vocalizations > non-vocalizations<br>Vocalization for positive on right (vocalizations > non-vocalizations<br>Valence for non-vocalizations on left (neutral > positive or negative) |

In *right R* the significant time window was 20-28 s post-stimulus onset. During this time window the triple interaction was driven by two double interactions (Table 2). First, the interaction Vocalization x Valence was significant for stimuli presented on the left (but not right or at the centre). Second, the interaction Vocalizations x Lateralization was significant for positive (but not neutral or negative) stimuli. These interactions were driven by the significant main effect of Vocalization for positive stimuli presented on the left, vocalizations yielding stronger activation than non-vocalizations. Post-hoc comparisons revealed during the same time window that among the vocalizations presented on the left positive ones yielded significantly greater activation than neutral or negative ones. Also positive vocalizations yielded significantly stronger activation when presented on the left than at the centre or on the right. Thus, taken together these results highlight the pro-eminence of positive vocalizations when presented on the left.

In *right L3* the significant time window was 20-28 s post-stimulus onset. During this time window the triple interaction was driven by three double interactions (Table 2). First, the interaction Vocalization x Valence was significant for stimuli presented on the left and on the right (but not at the centre). The latter was driven by a significant main effect of Vocalization on positive stimuli presented on the right, vocalizations yielding stronger activation than non-vocalizations. Second, the interaction Vocalization x Lateralization was significant for positive (but not neutral or negative) stimuli, driven by a significant main effect of vocalization on positive stimuli presented on the right or left (but not at the centre), vocalizations yielding stronger responses than non-vocalizations. Third, the interaction Valence x Lateralization was significant for vocalizations and for non-vocalizations. The latter was driven by a significant effect of Valence on non-vocalizations presented on the left; neutral non-vocalizations tended to yield stronger responses than positive or negative ones. Post-hoc comparisons revealed during the same time window that among the vocalizations presented on the left positive ones yielded significantly greater activation than negative ones. The same was the case among the vocalizations presented on the right, where positive ones yielded significantly greater activation than negative ones. Thus, taken together these results highlight the pro-eminence of positive vocalizations when presented on the left or on the right.

In summary, the results of the triple interaction and of the ensuing double interactions and main effects as well as the post-hoc comparisons highlight a significant pre-eminence of the left auditory space for the encoding of positive vocalizations in A1 and R bilaterally. In addition, left and right, but not central space is favoured for positive vocalizations in right L3.

***The encoding of emotional valence is modulated by the spatial origin of the sound*** *(hypothesis ii)*

The interaction Valence x Lateralization was significant bilaterally in VA. In the left hemisphere the significant time window was 10-14 s post-stimulus onset; post-hoc analysis did not yield any significant main effect of Valence at any position nor main effect of Lateralization on any valence (Table 3).

In the right hemisphere the interaction Valence x Lateralization was significant during 8-14 s plus 24-28 s. Post-hoc comparison showed that during the latter time window the main effect of valence was significant for sounds presented on the left side, negative sounds yielding lower activation than when presented centrally or on the right side. In summary, the spatial origin of the sound modulates the encoding of emotional valence within VA.

**Table 3**. Summary of significant double interaction Valence x Lateralization and the ensuing main effects in VA of the left and right hemispheres. For the time window of significant double interaction are listed the related main effects.

| **ROI with significant double interaction Valence x Lateralization (time window of significance)** | **Significant related main effect during the same time window** |
|---|---|
| ***Left hemisphere*** | |
| VA (10 - 14 s) | None |
| ***Right hemisphere*** | |
| VA (8 – 14 s) | None |
| VA (24 – 28 s) | Valence for sounds on left (positive > negative) |

***The spatial origin of the sound does not appear to impact differently the encoding of vocalizations vs. non-vocalizations*** *(hypothesis iii)*

The interaction Vocalization x Lateralization did not yield any significant results in either hemisphere.

***Spatial, emotional and vocalization selectivity***

A significant main effect of Lateralization was present in the left hemisphere in A1 (during the 10-14 s and 22-26 s time periods); in R (10-14 s and 18-36 s); and in M1 (10-14 s). The effect was driven by greater activation for contra- than ipsilateral stimuli.

### *Emotional valence modulates the encoding of vocalizations*

Significant interaction of Vocalization x Valence was present in either hemisphere. In the left hemisphere this was the case in A1 (12-26 s); R (14-18 s); L1 (10-18 s and 22-26); L2 (12-20 s and 24-28 s); M2 (22-26 s); and VA (4-16 s and 20-28 s). In the right hemisphere this was the case in A1 (14-18 s); R (12-28 s); L1 (14-24 s); L2 (10-26 s); L3 (12-18 s); L4 (14-18 s); M1 (14-18 s and 22-26 s); M3 (30-36 s); M4 (32-36 s); and VA (8-26 s). In A1, R, L1 and L2 the interactions appeared to be driven by the predominance of positive vocalizations and/or neutral non-vocalizations.

A significant main effect of Valence was present in several areas of either hemisphere. In the left hemisphere this was the case A1 (18-30 s); R (20-36 s); L1 (24-36 s); L2 (12-14 s and 18-36 s); L3 (6-12 s abd 16-36 s); M1 (16-24 s and 28-36 s); M2 (28-36 s); M4 (16-24 s and 28-36 s); and VA (6-28 s). In the right hemisphere it was the case in A1 (20-24 s); R (24-36 s); L2 (28-32 s); L3 (6-12 s and 24-36 s); M1 (20-24 s); M2 (18-22 s); M4 (16-24 s and 28-36 s); and VA (8-20 s). The effect tended to be driven by greater activation by vocalizations with positive rather than negative or neutral valence and by non-vocalizations with neutral rather than positive valence.

A significant main effect of Vocalization was present in either hemisphere. In the left hemisphere this was the case in L1 (6-22 s); L2 (6-26 s and 32-36 s); L3 (6-28 s and 32-36 s); M1 (8-12 s); and VA (6-28 s and 32-36 s). In the right hemisphere this was the case in L2 (6-28 s); L3 (6-26 s); and VA (6-28 s and 32-36 s). The effect was driven by greater activation by vocalizations than non-vocalizations.

## Discussion

Our results indicate that auditory spatial cues modulate the encoding of emotional valence in several early-stage auditory areas and in VA. The most striking effect is the pre-eminence of the left auditory space for the encoding of positive vocalizations. Furthermore, spatial cues appear to render emotional vocalizations more salient, as indicated by comparing our results with those of a previous study (Grisendi et al., 2019). The interactions of the type (human vocalizations vs. other environmental sounds), emotional valence and the spatial origin of the sound characterize the vocalization pathway within the early stage auditory areas and VA.

### *Pre-eminence of the left auditory space for positive vocalizations*

The left auditory space appears to favour the encoding of positive vocalizations within A1 and R in the left and right hemisphere (Fig. 3). In both hemispheres neural activity elicited by positive vocalizations presented on the left was higher that neural activity elicited by i) neutral or negative vocalizations presented at any of the three positions; or ii) non-vocalizations of any valence at any of the three positions. The bilateral involvement of A1 and R in favour of the left space speaks against a mere effect of contralateral space, which would yield stronger response to stimuli presented contralaterally.

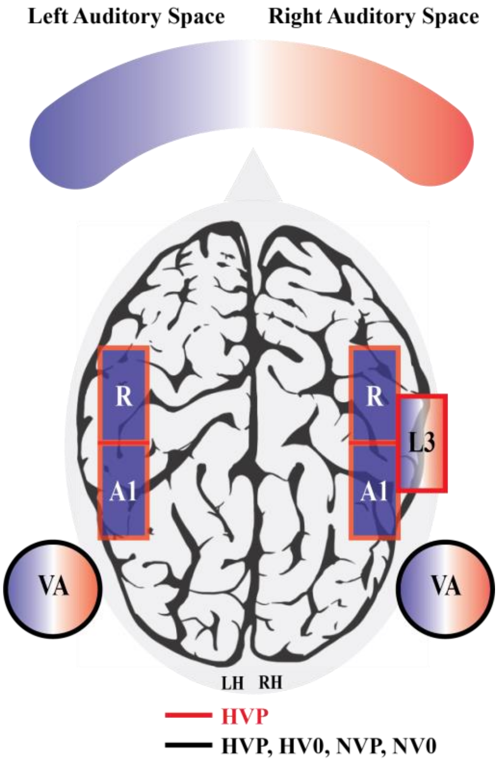

**Figure 3. Processing emotional value of the auditory space.** The left auditory space (in blue) is pre-eminent for positive vocalizations, positive vocalizations presented on the left side yielding greater activity in bilateral A1 and R than do neutral or negative vocalizations or any other stimuli at any of the three positions. To note that VA does not share the same preference for the left space. Spatial attributes modulate its activation by sound objects conveying positive or neutral emotional valence when presented on the right or left side (but not at the centre); this effect is not present for vocalizations alone. Left auditory space: in blue; right auditory space: in red; red lines: positive human vocalizations (HVP); black lines: positive and neutral human vocalizations and non-vocalizations.

The stronger encoding of positive vocalizations presented on the left side suggests that they may be more salient than when presented at other positions. The pre-eminence of the left auditory space, which we describe here, is reminiscent of the left-ear advantage, which was reported for emotional dichotic listening tasks in two studies (Erhan et al., 1998; Jäncke et al., 2001). Both studies compared emotional vs. neutral vocalizations, but did not discriminate between positive and negative valence. Their results have been interpreted in terms of right hemispheric competence for emotional processing (see also Gadea et al., 2011). Another series of studies used emotional valence of spoken words for spatial orienting of attention. Emotional word cues presented on the right side introduced spatial attentional bias for the following neutral sound (beep; Bertels et al., 2010). The interpretation of these results was influenced by the assumption that i) one-sided presentation of auditory stimuli is preferentially treated by the contralateral hemisphere and ii) the nature of the stimuli – verbal vs. emotional – tends to activate one hemisphere. Thus, the right side bias introduced by emotional words was eventually interpreted as prevailing influence of verbal content (Bertels et al., 2010). The nature of stimuli used in these studies, all verbal vocalizations, and the fact that they were presented mono-aurally, and not lateralized with interaural time (as here) or intensity differences, precludes their interpretation in terms of the emotional value of space.

The left-space preference, which we observed bilaterally in A1 and R, is greater for positive vocalizations than other stimuli. A similar preference for the contralateral, left space was reported by Kryklywy and colleagues (2013) in the right auditory cortex, with stronger responses to contralateral positive stimuli. Since the stimuli they used consisted to 75% of human vocalizations their finding is compatible with our results. It is to be noted that the preference for contralateral space is not limited to emotionally modulated sounds. Using emotionally neutral stimuli, a series of studies documented, at the level of the supratemporal plane, the preference for contralateral space, which together with right hemispheric dominance is a key feature of auditory spatial encoding (Deouell et al., 2007; Stecker et al., 2015; Derey et al., 2016 p.200, 2017; McLaughlin et al., 2016; Higgins et al., 2017; Da Costa et al., 2018).

Although compatible with evidence from previous studies, our results give a different picture of the emotional auditory space and its encoding within the early-stage auditory areas. We have documented a genuine pro-eminence of the left space for positive vocalizations and not simply a right hemispheric or contralateral dominance, the key observation being that left-sided positive vocalizations stand out within the primary auditory cortex of both hemispheres. The functional relevance of the emotional pro-eminence of the left auditory space needs to be investigated in future studies.

### *Spatial cues make emotional vocalizations more salient*

Two of our observations suggest that spatial cues render emotional vocalizations more salient. First, positive vocalizations presented on the right or the left yielded in right L3 greater activation than when presented centrally; vocalizations of other valence and non-vocalizations yielded smaller activation (Table 2). Second, the use of spatial cues appeared to enhance the discrimination of emotional valence in several early-stage areas. In a previous study, the same set of stimuli (human vocalizations and non-vocalizations of positive, neutral and negative valence), the same paradigm and an ANOVA based statistical analysis were used, albeit without lateralization (Grisendi et al., 2019). The juxtaposition of the distribution of significant interactions and significant main effects in early-stage areas and in VA highlights striking differences, which concern almost exclusively the factor Valence (and not Vocalization; Fig. 4). Main effect of Vocalization highlighted in both studies a very similar set of areas, with vocalizations yielding greater activation than non-vocalizations. Main effect of Valence was strikingly dissimilar, being significant in many more areas when spatial cues were used. The same was observed for the interaction Vocalization x Valence, with many more areas being significant when spatial cues were used; it is to be noted that in both studies the interaction was driven by greater responses to positive vocalizations. This increased saliency when spatial cues are used is not due to a modulation of emotional valence by lateralization; this interaction was only significant in VA but not in any of the early-stage areas.

The mechanisms by which spatial cues confer greater salience to emotional vocalizations is currently unknown. Interaural interactions during first cortical processing stages may enhance emotional stimuli, as does increasing intensity (Bach et al., 2008, 2009). Further study are needed to investigate whether the effect is associated uniquely with interaural time differences (used here) or whether interaural intensity differences or more complex spatial cues have the same effect.

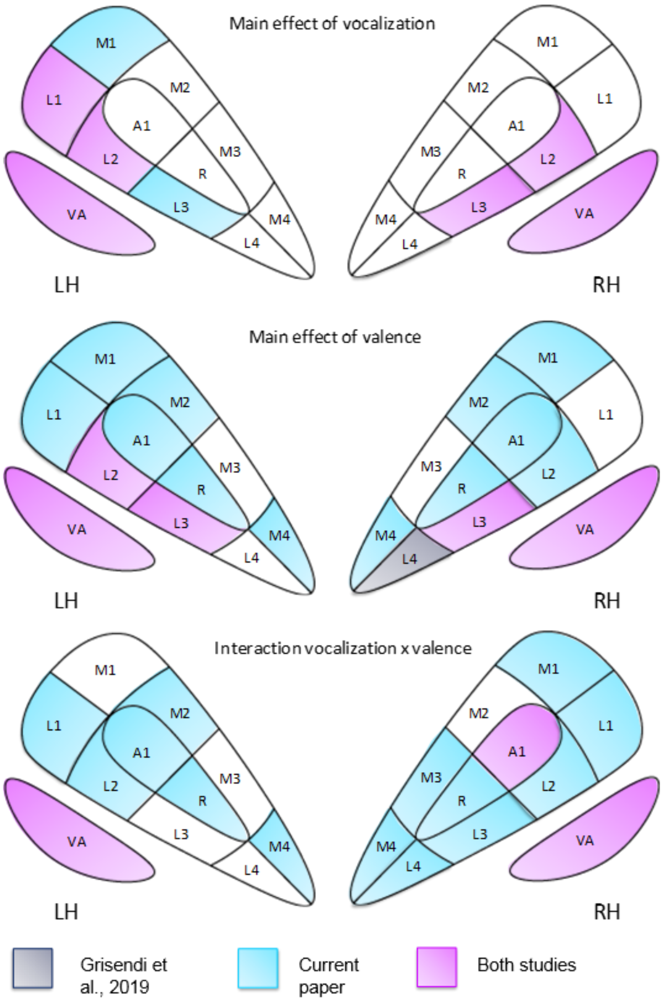

**Figure 4. Emotional sounds with or without spatial cues.** Juxtaposition of the results from the 2-way and 3-way ANOVAs found in the present (in light blue) and a previous study (Grisendi et al., 2019; in dark blue), which used the same set of stimuli, the same paradigm and an ANOVA based statistical approach. The former used lateralized stimuli, whereas the latter did not. In purple are indicated areas with a significant interaction or main effect in both studies. Whereas the main effect of Vocalizations highlights in both studies a very similar set of areas, the main effect of Valence and the interaction Vocalization x Valence yielded significant results in more areas when spatial cues where used.

### *Voice area: vocalizations are selectively modulated by emotional valence but not spatial cues*

Our analysis clearly showed that within VA the encoding of vocalizations is modulated by emotional valence, as did a series of previous studies (Belin et al., 2002; Grandjean et al., 2005; Ethofer et al., 2006, 2008, 2009, 2012; Beaucousin et al., 2007; Obleser et al., 2007, 2008; Bestelmeyer et al., 2017; Grisendi et al., 2019). The new finding is that this clear modulation of vocalizations by emotional valence is not paralleled by a modulation by the spatial origin of the sound. This is reminiscent of the findings of Kryklywy et al. (2013), who reported that emotional valence, but not spatial attributes, impacts the processing within the ventral stream. Their stimuli consisted to 75% of human vocalizations and may have driven the effect they observed.

In our study spatial information did not modulate significantly the encoding of vocalizations within VA. However, the spatial origin impacted the activity elicited by sound objects in general. Thus, positive and neutral sounds; i. e. vocalizations and non-vocalizations taken together, yielded stronger response than negative ones when presented on the left or on the right, as compared to a presentation at the centre. This preference for lateral space was present in both hemispheres.

## Conclusions

Previous behavioural studies (Erhan et al., 1998; Jäncke et al., 2001; Bertels et al., 2010) indicated that spatial origin impacts emotional processing of sounds, possibly via a preferential encoding of the contralateral space on the supratemporal plane (Kryklywy et al., 2013, 2018). We demonstrate here that there is a preference in terms of space, and not hemisphere, with a clear pre-eminence of the left auditory space for positive vocalizations. Positive vocalizations presented on the left side yield greater activity in bilateral A1 and R than do neutral or negative vocalizations or any other stimuli at any of the three positions. VA does not share the same preference for the left space. Spatial attributes modulate its activation by sound objects conveying positive or neutral emotional valence when presented on the right or left side (but not at the centre); this effect is not present for vocalizations alone. Comparison with a previous study (Grisendi et al., 2019) indicates that spatial cues may render emotional valence more salient within the early-stage auditory areas.

## Ethical approval

All procedures performed were in accordance and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards and ethical approval was obtained from the Ethical Committee of the Canton de Vaud (reference number 282/08).

## References

Aeschlimann M, Knebel J-F, Murray MM, Clarke S (2008) Emotional Pre-eminence of Human Vocalizations. Brain Topogr 20:239–248.

Ahveninen J, Jääskeläinen IP, Raij T, Bonmassar G, Devore S, Hämäläinen M, Levänen S, Lin F-H, Sams M, Shinn-Cunningham BG, Witzel T, Belliveau JW (2006) Task-modulated “what” and “where” pathways in human auditory cortex. Proc Natl Acad Sci 103:14608–14613.

Anourova I, Nikouline VV, Ilmoniemi RJ, Hotta J, Aronen HJ, Carlson S (2001) Evidence for Dissociation of Spatial and Nonspatial Auditory Information Processing. NeuroImage 14:1268–1277.

Arnal LH, Poeppel D, Giraud A-L (2015) Temporal coding in the auditory cortex. Handb Clin Neurol 129:85–98.

Arnott SR, Binns MA, Grady CL, Alain C (2004) Assessing the auditory dual-pathway model in humans. NeuroImage 22:401–408.

Bach DR, Neuhoff JG, Perrig W, Seifritz E (2009) Looming sounds as warning signals: the function of motion cues. Int J Psychophysiol Off J Int Organ Psychophysiol 74:28–33.

Bach DR, Schächinger H, Neuhoff JG, Esposito F, Di Salle F, Lehmann C, Herdener M, Scheffler K, Seifritz E (2008) Rising sound intensity: an intrinsic warning cue activating the amygdala. Cereb Cortex N Y N 1991 18:145–150.

Beaucousin V, Lacheret A, Turbelin M-R, Morel M, Mazoyer B, Tzourio-Mazoyer N (2007) FMRI Study of Emotional Speech Comprehension. Cereb Cortex 17:339–352.

Belin P, Zatorre RJ, Ahad P (2002) Human temporal-lobe response to vocal sounds. Cogn Brain Res 13:17–26.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403:309–312.

Bertels J, Kolinsky R, Morais J (2010) Emotional valence of spoken words influences the spatial orienting of attention. Acta Psychol (Amst) 134:264–278.

Besle J, Mougin O, Sánchez-Panchuelo R-M, Lanting C, Gowland P, Bowtell R, Francis S, Krumbholz K (2019) Is Human Auditory Cortex Organization Compatible With the Monkey Model? Contrary Evidence From Ultra-High-Field Functional and Structural MRI. Cereb Cortex N Y N 1991 29:410–428.

Bestelmeyer PEG, Kotz SA, Belin P (2017) Effects of emotional valence and arousal on the voice perception network. Soc Cogn Affect Neurosci Available at: https://academic.oup.com/scan/article/doi/10.1093/scan/nsx059/3753554/Effects-of-emotional-valence-and-arousal-on-the [Accessed May 30, 2017].

Bourquin NM-P, Murray MM, Clarke S (2013) Location-independent and location-linked representations of sound objects. NeuroImage 73:40–49.

Chiry O, Tardif E, Magistretti PJ, Clarke S (2003) Patterns of calcium-binding proteins support parallel and hierarchical organization of human auditory areas. Eur J Neurosci 17:397–410.

Clarke S, Bellmann A, Meuli RA, Assal G, Steck AJ (2000) Auditory agnosia and auditory spatial deficits following left hemispheric lesions: evidence for distinct processing pathways. Neuropsychologia 38:797–807.

Clarke S, Bellmann Thiran A, Maeder P, Adriani M, Vernet O, Regli L, Cuisenaire O, Thiran J-P (2002) What and where in human audition: selective deficits following focal hemispheric lesions. Exp Brain Res Exp Hirnforsch Expérimentation Cérébrale 147:8–15.

Clarke S, Geiser E (2015) Roaring lions and chirruping lemurs: How the brain encodes sound objects in space. Neuropsychologia 75:304–313.

Clarke S, Morosan P (2012) Architecture, connectivity and transmitter receptors of human auditory cortex. In: In Poeppel, D., Overath, T., Popper, A. N., and Fay, R.R. (eds). (2012). Human Auditory Cortex., Springer Science+Business Media, LLC, New York.

Clarke S, Rivier F (1998) Compartments within human primary auditory cortex: evidence from cytochrome oxidase and acetylcholinesterase staining. Eur J Neurosci 10:741–745.

Courtois R, Petot J-M, Lignier B, Lecocq G, Plaisant O (2018) [Does the French Big Five Inventory evaluate facets other than the Big Five factors?]. L'Encephale 44:208–214.

Da Costa S, Bourquin NM-P, Knebel J-F, Saenz M, van der Zwaag W, Clarke S (2015) Representation of Sound Objects within Early-Stage Auditory Areas: A Repetition Effect Study Using 7T fMRI. PloS One 10:e0124072.

Da Costa S, Clarke S, Crottaz-Herbette S (2018) Keeping track of sound objects in space: The contribution of early-stage auditory areas. Hear Res Available at: http://www.sciencedirect.com/science/article/pii/S0378595517306032 [Accessed July 9, 2018].

Da Costa S, Saenz M, Clarke S, Zwaag W van der (2014) Tonotopic Gradients in Human Primary Auditory Cortex: Concurring Evidence From High-Resolution 7 T and 3 T fMRI. Brain Topogr:1–4.

Da Costa S, van der Zwaag W, Marques JP, Frackowiak RSJ, Clarke S, Saenz M (2011) Human Primary Auditory Cortex Follows the Shape of Heschl's Gyrus. J Neurosci 31:14067–14075.

Da Costa S, van der Zwaag W, Miller LM, Clarke S, Saenz M (2013) Tuning In to Sound: Frequency-Selective Attentional Filter in Human Primary Auditory Cortex. J Neurosci 33:1858–1863.

De Meo R, Bourquin NM-P, Knebel J-F, Murray MM, Clarke S (2015) From bird to sparrow: Learning-induced modulations in fine-grained semantic discrimination. NeuroImage 118:163–173.

De Santis L, Clarke S, Murray MM (2007a) Automatic and intrinsic auditory “what” and “where” processing in humans revealed by electrical neuroimaging. Cereb Cortex N Y N 1991 17:9–17.

De Santis L, Spierer L, Clarke S, Murray MM (2007b) Getting in touch: segregated somatosensory what and where pathways in humans revealed by electrical neuroimaging. NeuroImage 37:890–903.

Deouell LY, Heller AS, Malach R, D’Esposito M, Knight RT (2007) Cerebral Responses to Change in Spatial Location of Unattended Sounds. Neuron 55:985–996.

Derey K, Rauschecker JP, Formisano E, Valente G, de Gelder B (2017) Localization of complex sounds is modulated by behavioral relevance and sound category. J Acoust Soc Am 142:1757.

Derey K, Valente G, de Gelder B, Formisano E (2016) Opponent Coding of Sound Location (Azimuth) in Planum Temporale is Robust to Sound-Level Variations. Cereb Cortex N Y N 1991 26:450–464.

Duffour-Nikolov C, Tardif E, Maeder P, Thiran AB, Bloch J, Frischknecht R, Clarke S (2012) Auditory spatial deficits following hemispheric lesions: Dissociation of explicit and implicit processing. Neuropsychol Rehabil 22:674–696.

Eramudugolla R, McAnally KI, Martin RL, Irvine DRF, Mattingley JB (2008) The role of spatial location in auditory search. Hear Res 238:139–146.

Erhan H, Borod JC, Tenke CE, Bruder GE (1998) Identification of emotion in a dichotic listening task: event-related brain potential and behavioral findings. Brain Cogn 37:286–307.

Ethofer T, Anders S, Wiethoff S, Erb M, Herbert C, Saur R, Grodd W, Wildgruber D (2006) Effects of prosodic emotional intensity on activation of associative auditory cortex: NeuroReport 17:249–253.

Ethofer T, Bretscher J, Gschwind M, Kreifelts B, Wildgruber D, Vuilleumier P (2012) Emotional voice areas: anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. Cereb Cortex N Y N 1991 22:191–200.

Ethofer T, Kreifelts B, Wiethoff S, Wolf J, Grodd W, Vuilleumier P, Wildgruber D (2008) Differential Influences of Emotion, Task, and Novelty on Brain Regions Underlying the Processing of Speech Melody. J Cogn Neurosci 21:1255–1268.

Ethofer T, Van De Ville D, Scherer K, Vuilleumier P (2009) Decoding of Emotional Information in Voice-Sensitive Cortices. Curr Biol 19:1028–1033.

Formisano E, Kim D-S, Salle FD, Moortele P-F van de, Ugurbil K, Goebel R (2003) Mirror-Symmetric Tonotopic Maps in Human Primary Auditory Cortex. Neuron 40:859–869.

Gadea M, Espert R, Salvador A, Martí-Bonmatí L (2011) The sad, the angry, and the asymmetrical brain: dichotic listening studies of negative affect and depression. Brain Cogn 76:294–299.

Goedhart AD, Van Der Sluis S, Houtveen JH, Willemsen G, De Geus EJC (2007) Comparison of time and frequency domain measures of RSA in ambulatory recordings. Psychophysiology 44:203–215.

Grandjean D, Sander D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P (2005) The voices of wrath: brain responses to angry prosody in meaningless speech. Nat Neurosci 8:145–146.

Grisendi T, Reynaud O, Clarke S, Da Costa S (2019) Processing pathways for emotional vocalizations. Brain Struct Funct.; In revision.

Hackett TA, Preuss TM, Kaas JH (2001) Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. J Comp Neurol 441:197–222.

Hart HC, Palmer AR, Hall DA (2004) Different areas of human non-primary auditory cortex are activated by sounds with spatial and nonspatial properties. Hum Brain Mapp 21:178–190.

Higgins NC, McLaughlin SA, Da Costa S, Stecker GC (2017) Sensitivity to an Illusion of Sound Location in Human Auditory Cortex. Front Syst Neurosci 11:35.

Jäncke L, Buchanan TW, Lutz K, Shah NJ (2001) Focused and nonfocused attention in verbal and emotional dichotic listening: an FMRI study. Brain Lang 78:349–363.

Kasper L, Bollmann S, Diaconescu AO, Hutton C, Heinzle J, Iglesias S, Hauser TU, Sebold M, Manjaly Z-M, Pruessmann KP, Stephan KE (2017) The PhysIO Toolbox for Modeling Physiological Noise in fMRI Data. J Neurosci Methods 276:56–72.

Kryklywy JH, Macpherson EA, Greening SG, Mitchell DGV (2013) Emotion modulates activity in the ‘what’ but not ‘where’ auditory processing pathway. NeuroImage 82:295–305.

Kryklywy JH, Macpherson EA, Mitchell DGV (2018) Decoding auditory spatial and emotional information encoding using multivariate versus univariate techniques. Exp Brain Res:1–9.

Lavan N, Rankin G, Lorking N, Scott S, McGettigan C (2017) Neural correlates of the affective properties of spontaneous and volitional laughter types. Neuropsychologia 95:30–39.

Leitman DI, Wolf DH, Ragland JD, Laukka P, Loughead J, Valdez JN, Javitt DC, Turetsky BI, Gur RC (2010) “It’s Not What You Say, But How You Say it”: A Reciprocal Temporo-frontal Network for Affective Prosody. Front Hum Neurosci 4:19.

Maeder PP, Meuli RA, Adriani M, Bellmann A, Fornari E, Thiran JP, Pittet A, Clarke S (2001) Distinct pathways involved in sound recognition and localization: a human fMRI study. NeuroImage 14:802–816.

Marques JP, Kober T, Krueger G, van der Zwaag W, Van de Moortele P-F, Gruetter R (2010) MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. NeuroImage 49:1271–1281.

McLaughlin SA, Higgins NC, Stecker GC (2016) Tuning to Binaural Cues in Human Auditory Cortex. J Assoc Res Otolaryngol JARO 17:37–53.

Moerel M, De Martino F, Formisano E (2014) An anatomical and functional topography of human auditory cortical areas. Front Neurosci 8:225.

Obleser J, Eisner F, Kotz SA (2008) Bilateral Speech Comprehension Reflects Differential Sensitivity to Spectral and Temporal Features. J Neurosci 28:8116–8123.

Obleser J, Zimmermann J, Van Meter J, Rauschecker JP (2007) Multiple Stages of Auditory Speech Perception Reflected in Event-Related fMRI. Cereb Cortex 17:2251–2257.

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–113.

Pernet CR, McAleer P, Latinus M, Gorgolewski KJ, Charest I, Bestelmeyer PEG, Watson RH, Fleming D, Crabbe F, Valdes-Sosa M, Belin P (2015) The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. NeuroImage 119:164–174.

Rademacher J, Morosan P, Schleicher A, Freund HJ, Zilles K (2001) Human primary auditory cortex in women and men. Neuroreport 12:1561–1565.

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12:718–724.

Rey B, Frischknecht R, Maeder P, Clarke S (2007) Patterns of recovery following focal hemispheric lesions: relationship between lasting deficit and damage to specialized networks. Restor Neurol Neurosci 25:285–294.

Rivier F, Clarke S (1997) Cytochrome Oxidase, Acetylcholinesterase, and NADPH-Diaphorase Staining in Human Supratemporal and Insular Cortex: Evidence for Multiple Auditory Areas. NeuroImage 6:288–304.

Stecker GC, McLaughlin SA, Higgins NC (2015) Monaural and binaural contributions to interaural-level-difference sensitivity in human auditory cortex. NeuroImage 120:456–466.

Tajadura-Jiménez A, Larsson P, Väljamäe A, Västfjäll D, Kleiner M (2010a) When room size matters: acoustic influences on emotional responses to sounds. Emot Wash DC 10:416–422.

Tajadura-Jiménez A, Väljamäe A, Asutay E, Västfjäll D (2010b) Embodied auditory perception: the emotional impact of approaching and receding sound sources. Emot Wash DC 10:216–229.

Thiran AB, Clarke S (2003) Preserved use of spatial cues for sound segregation in a case of spatial deafness. Neuropsychologia 41:1254–1261.

Tissieres I, Crottaz-Herbette S, Clarke S (2019) Implicit representation of the auditory space: contribution of the left and right hemispheres. Brain Struct Funct.

van der Zwaag W, Gentile G, Gruetter R, Spierer L, Clarke S (2011) Where sound position influences sound object representations: A 7-T fMRI study. NeuroImage 54:1803–1811.

Viceic D, Fornari E, Thiran J-P, Maeder PP, Meuli R, Adriani M, Clarke S (2006) Human auditory belt areas specialized in sound recognition: a functional magnetic resonance imaging study. Neuroreport 17:1659–1662.

Wallace MN, Johnston PW, Palmer AR (2002) Histochemical identification of cortical areas in the auditory region of the human brain. Exp Brain Res 143:499–508.

Wildgruber D, Riecker A, Hertrich I, Erb M, Grodd W, Ethofer T, Ackermann H (2005) Identification of emotional intonation evaluated by fMRI. NeuroImage 24:1233–1241.

Zigmond AS, Snaith RP (1983) The hospital anxiety and depression scale. Acta Psychiatr Scand 67:361–370.

Zilles K, Armstrong E, Schleicher A, Kretschmann HJ (1988) The human pattern of gyrification in the cerebral cortex. Anat Embryol (Berl) 179:173–179.

# STUDY C: Functional correlations between early-stage auditory areas and the voice area

Tiffany Grisendi, Stephanie Clarke and Sandra Da Costa

In preparation

# Functional correlations between early-stage auditory areas and the voice area

**Tiffany Grisendi[1], Stephanie Clarke[1] and Sandra Da Costa[2]**

[1] Service de Neuropsychologie et de Neuroréhabilitation, Centre Hospitalier Universitaire Vaudois (CHUV) and University of Lausanne, 1011 Lausanne, Switzerland

[2] Centre d'Imagerie Biomédicale, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

**Corresponding author:**
Dr. Sandra Da Costa
EPFL SB CIBM
CH F0 562 – Station 6
CH-1015 Lausanne
Tel.: +41 21 693 76 87
ORCID 0000-0002-8641-0494
Email: sandra.dacosta@epfl.ch

Number of pages: 28
Number of figures: 10
Number of tables: 2
Number of words in Abstract: 103

**Acknowledgements:** This work was supported by the Swiss National Science Foundation Grant to S. Clarke (FNS 320030-159708) and by the Centre d'Imagerie BioMédicale (CIBM) of the UNIL, UNIGE, HUG, CHUV, EPFL and the Leenaards and Jeantet Foundations.

**Conflict of interest:** The authors declare no competing financial interests.

**Sharing data:** The dataset of this manuscript is available upon request to the corresponding author.

## Abstract

Human vocalizations processing and their modulation by emotional valence and/or localization involves brain regions, such as voice area (VA) and early-stage auditory areas. Using two separate datasets acquired at 7T fMRI, we have investigated the functional correlations between early-stage auditory areas and VA as modulated by the category of sound, valence and position. The functional correlations between VA, primary auditory cortex (PAC) and lateral belt areas were strongly modulated by the category of sound, weakly by the spatial positioning and not by the affective content. Human vocalizations produce stronger functional correlations between VA, PAC and lateral belt areas, compared to non-vocal environmental sounds.

## Keywords

**Human vocalizations, emotions, auditory belt areas, voice area, lateralization, 7T fMRI**

## Abbreviations

| | |
|---|---|
| AI | primary auditory area |
| HVN | human vocalizations with negative emotional valence |
| HVP | human vocalizations with positive emotional valence |
| HV0 | human vocalizations with neutral emotional valence |
| NVN | non-vocalizations with negative emotional valence |
| NVP | non-vocalizations with positive emotional valence |
| NV0 | non-vocalizations with neutral emotional valence |
| PAC | primary auditory cortex |
| R | rostral (primary) auditory area |
| VA | voice area |

## Introduction

Auditory information, and more specifically emotional sounds, is crucial in our everyday life. The voice area (VA) and early-stage auditory areas are part of networks responsible for the processing of the emotional valence of human vocalizations, and that it is based on separated sensory, emotional and cognitive processes, as described by Frühholz et al., (2016) and Schirmer and Kotz, (2006). A core network of cortical and subcortical areas (e.g., medial frontal cortex, inferior frontal gyrus, insula, auditory cortex, superior temporal cortex, basal ganglia, amygdala and cerebellum), is responsible for the processing of affective auditory information in general, and additional regions (e.g., hippocampus, thalamus, inferior colliculus) complete this core network for the processing of more specific categories of emotional sounds.

Anatomical, histological and functional studies described early-stage auditory areas located on the supratemporal plane (Clarke and Morosan, 2012). Surrounding the primary auditory cortex (PAC), non-primary areas were identified histologically in *post-mortem* brains (Chiry et al., 2003; Clarke and Rivier, 1998; Hackett et al., 2001; Rivier and Clarke, 1997; Wallace et al., 2002). Definition of these areas for activation studies was performed based on their Talairach coordinates (van der Zwaag et al., 2011; Viceic et al., 2006), in combination with the definition of PAC by tonotopic mapping (Da Costa et al., 2015, 2018). The latter studies identified in total 10 areas: two primary, A1 and R, located on Heschl's gyrus using as landmark the two mirror-reversed tonotopic maps (Da Costa et al., 2011); four areas on the lateral part , designated as L1, L2, L3 and L4 (from posterior to anterior); and four areas medially, designated M1, M2, M3 and M4 (see e.g. Da Costa et al., 2018).

VA is a region located in the middle part of the superior temporal gyrus/sulcus that responds more strongly to human vocalizations than to any other non-vocal sounds (Belin et al., 2000). This region is not only selective for human vocalizations but is also sensitive to the emotional valence of the vocalizations (Beaucousin et al., 2007; Belin et al., 2002; Bestelmeyer et al., 2017; Ethofer et al., 2009, 2006; Grandjean et al., 2005; Obleser et al., 2007). In addition to VA, the auditory cortex is modulated by the emotional value of vocalizations (Arnal et al., 2015; Lavan et al., 2017; Leitman et al., 2010; Wildgruber et al., 2004). More specifically, lateral early-stage auditory areas (L1, L2 and L3) present stronger responses to human vocalizations compared to non-vocalizations and/or these responses were dependent on the emotional value of the stimuli. On the other hand, emotional valence modulated specifically the responses of VA to human vocalizations and not to non-vocalizations (Grisendi et al., 2019a). Finally, the processing of vocalizations by VA is modulated by the emotional valence, but not by the lateralization of the sounds (Grisendi et al., 2019b).

In a couple of recent studies, Grisendi et al., (2019a,b) use the advantages of ultra-high field fMRI at 7 Tesla to identify individual early-stage auditory areas. Their results showed that neural responses within the lateral belt areas L1, L2 and L3 are modulated by vocalizations and/or emotional content. These areas are also known to be modulated by spatial cues (Da Costa et al., 2018). Functional correlations between L1, L2, L3 and VA were observed for different emotional valences (positive, neutral, negative) and for different sound categories (human vocalizations and environmental sounds; Grisendi et al., 2019a). Correlation of the BOLD signal between VA and AMY in the right hemisphere was observed for positive human vocalizations only. This suggests that the emotional information contained in human vocalizations is conveyed to VA via a dual input, in one side the early-stage auditory areas L1, L2 and L3 and on the other side the amygdala.

In the present study, we investigated the functional correlation between VA and early-stage auditory areas, and their modulation by vocalization, valence and lateralization of the sounds. Based on previous findings, we expected (1) VA to be functionally correlated with early-stage auditory areas; and (2) these correlations to concern specific stimuli and conditions. These hypotheses were tested by looking at correlations between the BOLD signals from the above regions of interest (ROIs) for different conditions of vocalization, valence and lateralization.

## Materials and Methods

We analyse here datasets from two previous studies Grisendi et al., (2019a,b).

### *Study 1: Processing of emotions within AC (Grisendi et al., 2019a)*

In summary, eleven healthy, normal-hearing and French-speaking subjects participated in the study, and the dataset of one subject was discarded due to data acquisition problems (total number of subjects = 10). The experimental design consisted of one fMRI session (~60 min) during which the participants listened passively to human vocalizations and environmental sounds with positive, neutral or negative emotional valence. Auditory stimuli (Aeschlimann et al., 2008) were presented in blocks of eleven different sounds from the same category, with the same emotional valence. The experiment was composed of three runs, in which blocks were pseudo-randomized. Each fMRI run last 8 minutes, with first a 30-s silent “rest” condition, followed by 14 blocks of 30 s (22 s of sounds + 8s of silence), and finally a 30-s silent “rest” condition.

### *Study 2: Processing of emotions and lateralization within AC (Grisendi et al., 2019b)*

In this study, fourteen healthy, normal-hearing and French-speaking subjects participated in the study, and the dataset of one subject was discarded due to data acquisition problems (total number of subjects = 13). The experimental design consisted of two fMRI sessions (~60 min each) during which the participants listened passively to human vocalizations and environmental sounds with positive, neutral or negative emotional valence and lateralized to the left (-60°; ITD = -0.3 ms), to the centre (0°) or to the right (+60°; ITD = +0.3ms). Auditory stimuli were from the same battery as for study 1, and were presented in blocks of eleven different sounds from the same category, with the same emotional valence and the same lateralization. The experiment was composed of twelve runs, in which blocks were pseudo-randomized. Each fMRI run last 6 minutes, with first a 20-s silent "rest" condition, followed by 9 blocks of 36 s (22 s of sounds + 14s of silence), and finally a 20-s silent "rest" condition.

### *MRI acquisitions*

MRI data acquisition was performed on a 7-Tesla MRI scanner (Siemens MAGNETOM scanner, Siemens Medical Solutions) with a 8-channel head rf-coil, and a 32-channel head rf-coil for the first and second study, respectively.

### *ROIs definition*

In addition to the emotional sounds runs, a tonotopic mapping paradigm (Da Costa et al., 2011) and a voice localizer (Pernet et al., 2015; only for study 2) were implemented to define our ROIs. The tonotopic mapping was used to identify individual primary auditory areas (A1 and R) and the belt areas (L1, L2, L3, L4, M1, M2, M3 and M4), as previously described in Da Costa et al., (2015, 2018), in both studies. VA was defined in different ways between the two studies. In the first study, the contrast "human vocalizations *vs.* non-vocalizations" was used to identify the VA of each subject, while in the second study the voice localizer was used to identify this region. The voice localizer was composed of human vocalizations and non-vocalizations from a battery different from the one used in the main experiment.

### *Correlation analysis*

For both datasets, individual BOLD time courses for each ROI were extracted using BrainVoyager (BrainVoyager v20.6, Brain Innovation, Maastricht, Netherlands) and analysed into MATLAB (R2015b, The MathWorks, Inc., Natick, Massachusetts, United States). Correlations between the BOLD signals of VA, A1 and R, and belt auditory areas (L1, L2, L3, L4, M1, M2, M3 and M4) were performed for human vocalizations and non-vocalizations with modulation of valence (negative, neutral and positive) and lateralization (left, centre and right). Only significant correlations ($p<0.01$ and adjusted $R^2 >0.6$) were taken into account for further analyses. These significant correlations were represented in heatmaps illustrating the adjusted-$R^2$ correlation factor. Heatmaps were used only to descriptive purpose, as a visual representation of the strength of functional correlations between two ROIs in a color-coded manner. We finally performed two separate ANOVAs for each study on the adjusted $R^2$ correlation factor computed between VA, A1 and R, the lateral auditory areas (L1, L2, L3 and L4) and the medial auditory areas (M1, M2, M3 and M4): a 2-way ANOVA (for study 1) with Vocalization (human-vocalizations, non-vocalizations) x Valence (negative, neutral, positive); and a 3-way ANOVA (for study 2) with Vocalization (human-vocalizations, non-vocalizations) x Valence (negative, neutral, positive) x Lateralization (left, centre, right).

## Results

### *Study 1: Processing of emotions within AC*

In the first study (Grisendi et al., 2019a), the heatmaps (Figure 1-2) indicated that the correlations between the different regions were stronger in the right hemisphere than in the left hemisphere, as well as for the non-vocalizations compared to the human vocalizations. The strongest correlations were found between the regions of A1 and R and the lateral belt areas (L1, L2 and L3 especially), as well as with VA but only for the human vocalizations. The region M4 is less correlated with the other regions in the left hemisphere compared to the right one. There is also a clear dissociation between the functional correlations of the lateral and medial belt areas. Indeed, A1, R and VA, for human vocalizations only, are strongly correlated with the lateral regions, but only sparsely correlated with the medial belt areas.

The 2-way ANOVA: Vocalization (human vocalizations, non-vocalizations) x Valence (negative, neutral, positive) revealed a main effect of vocalization in the left hemisphere between VA, A1 and R, and lateral belt areas (L1, L2 and L3), and between VA and M1 in the right hemisphere (Table 1 and 2). This effect is due to stronger correlations between VA and early-stage auditory areas for human vocalizations compared to non-vocalizations (Figure 9 and 10). A main effect of valence was found between VA and A1 in the left hemisphere, and between VA and M1 in the right hemisphere. The former might be driven by stronger correlations between VA and A1 for the positive and neutral stimuli compared to the negative ones.

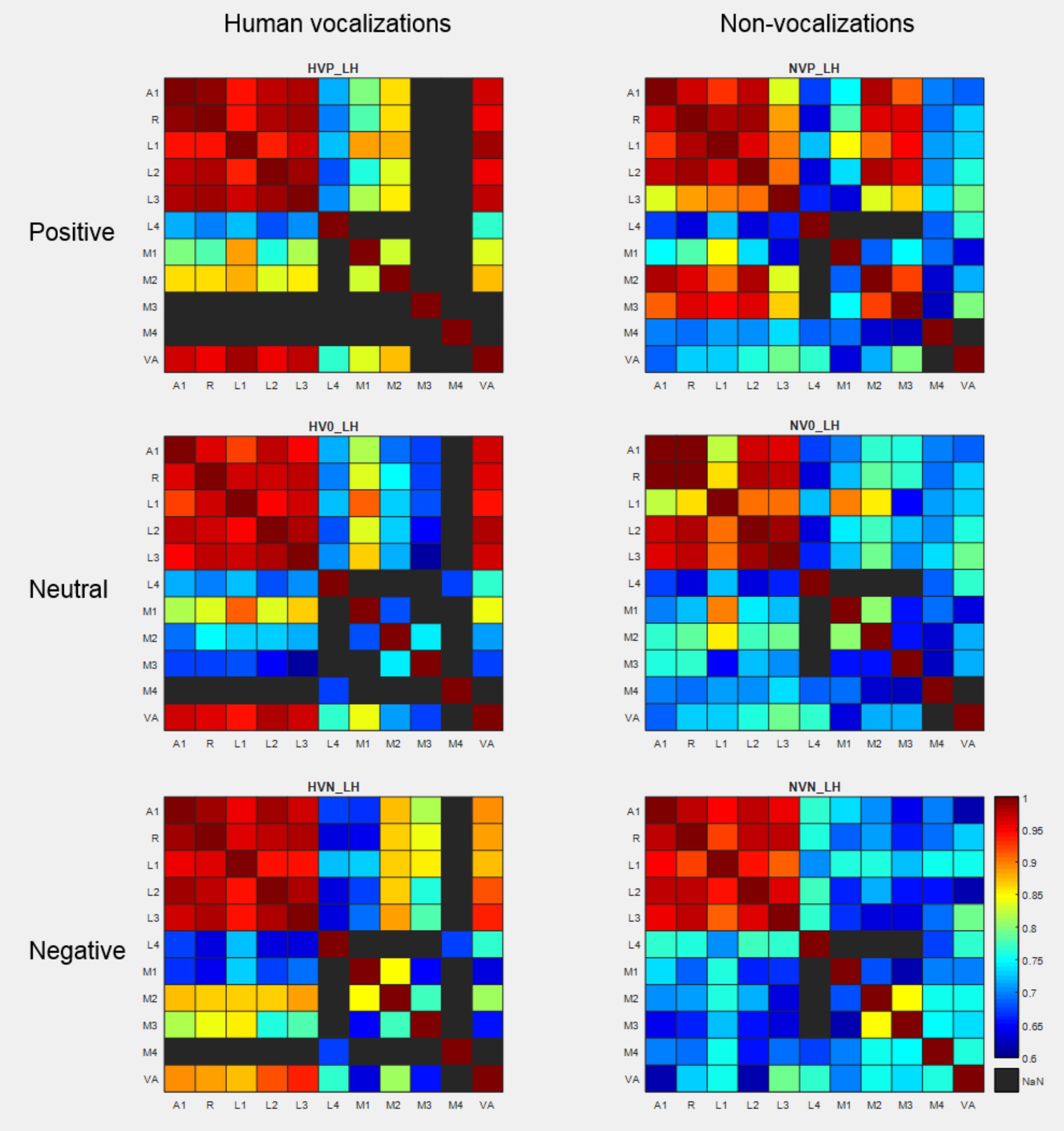

**Figure 1: Heatmaps study 1 left hemisphere**. Heatmaps for the data of the study 1 representing the correlations (adjusted $R^2$) between the BOLD signals of early-stage auditory areas A1, R, L1, L2, L3, L4, M1, M2, M3 and M4 and VA during the presentation of human vocalizations (left column) and non-vocalizations (right column) with negative, neutral and positive valence (top, middle and bottom rows, respectively). The color code represents the strength of the correlation (between [0.6 – 1]).The NaN values represent correlations that are statistically non-significant ($p > 0.01$ and/or adjusted $R^2 < 0.6$).

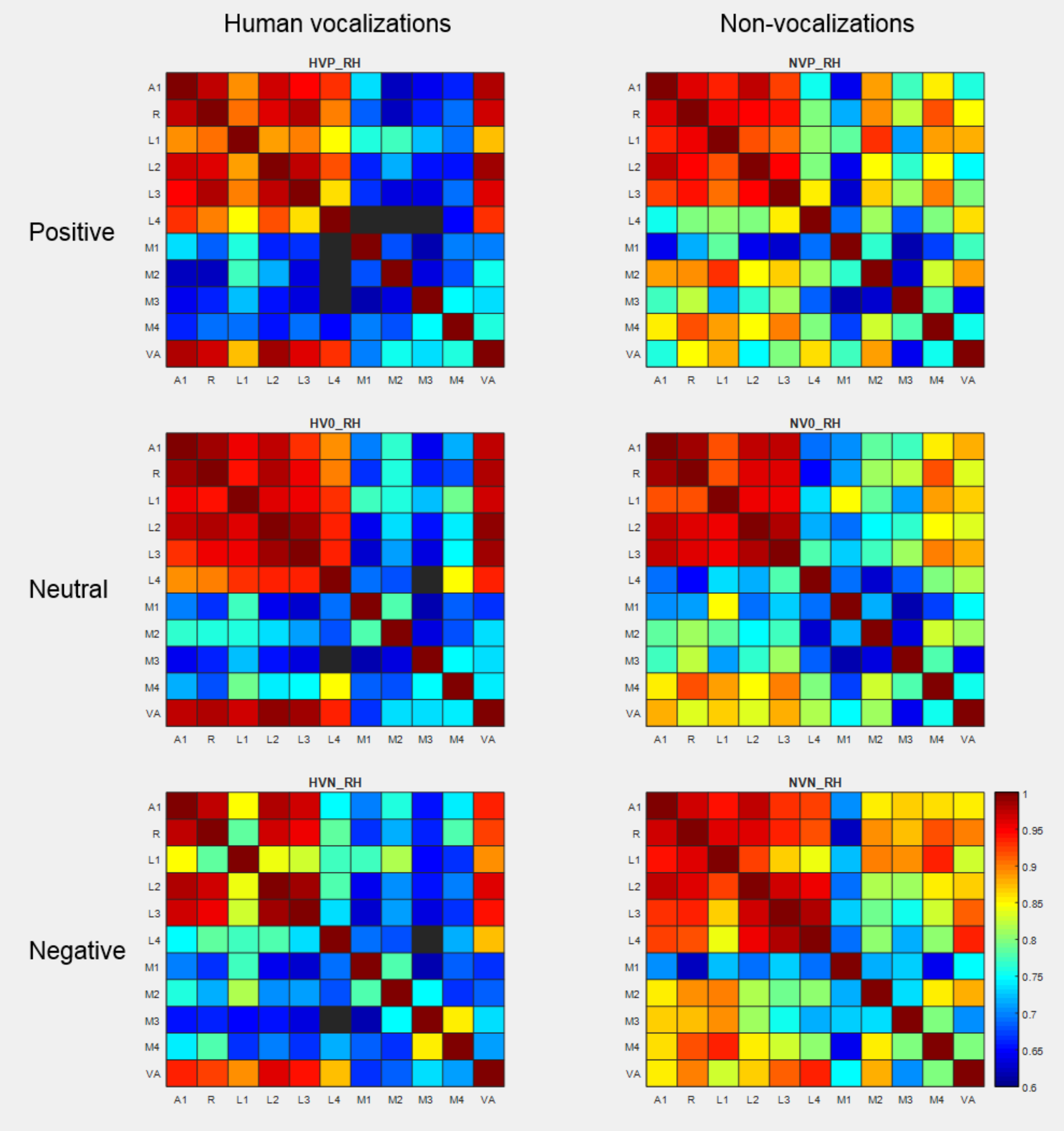

**Figure 2: Heatmaps study 1 right hemisphere**. Heatmaps for the data of the study 1 representing the correlations (adjusted $R^2$) between the BOLD signals of early-stage auditory areas A1, R, L1, L2, L3, L4, M1, M2, M3 and M4 and VA during the presentation of human vocalizations (left column) and non-vocalizations (right column) with negative, neutral and positive valence (top, middle and bottom rows, respectively). The color code represents the strength of the correlation (between [0.6 – 1]).The NaN values represent correlations that are statistically non-significant ($p > 0.01$ and/or adjusted $R^2 < 0.6$).

| LH | Study | A1 | R | L1 | L2 | L3 | L4 | M1 | M2 | M3 | M4 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Main effect of voc** | a | 0.0001 | 0.0139 | 0.0401 | 0.0182 | 0.0054 | 1.0000 | 0.3229 | 0.2724 | 0.3357 | 0.4226 |
| | b | 0.0017 | 0.0012 | 0.0043 | 0.0030 | 0.0044 | 0.0094 | 0.0011 | 0.0028 | 0.0012 | 0.5165 |
| **Main effect of val** | a | 0.0016 | 0.5000 | 0.7137 | 0.2072 | 0.5000 | 1.0000 | 0.7775 | 0.4666 | 0.6101 | 0.5000 |
| | b | 0.2241 | 0.2950 | 0.1510 | 0.3589 | 0.4608 | 0.8128 | 0.1188 | 0.1315 | 0.0686 | 0.3348 |
| **Main effect of lat** | a | | | | | | | | | | |
| | b | 0.6277 | 0.6012 | 0.4797 | 0.6538 | 0.3678 | 0.7762 | 0.7476 | 0.4500 | 0.4767 | 0.3644 |
| **Inter voc x val** | a | 0.1256 | 0.3125 | 0.4821 | 0.4129 | 0.4571 | 1.0000 | 0.3985 | 0.3496 | 0.7453 | 0.3219 |
| | b | 0.2071 | 0.2962 | 0.3148 | 0.3804 | 0.4425 | 0.5276 | 0.1056 | 0.1118 | 0.0986 | 0.2455 |
| **Inter voc x lat** | a | | | | | | | | | | |
| | b | 0.5389 | 0.5285 | 0.9759 | 0.6162 | 0.3498 | 0.6045 | 0.9500 | 0.5723 | 0.3028 | 0.0742 |
| **Inter val x lat** | a | | | | | | | | | | |
| | b | 0.3275 | 0.3494 | 0.4294 | 0.3845 | 0.3770 | 0.6396 | 0.6127 | 0.5516 | 0.5576 | 0.9144 |
| **Inter voc x val x lat** | a | | | | | | | | | | |
| | b | 0.6512 | 0.6423 | 0.7541 | 0.6689 | 0.5749 | 0.6412 | 0.6012 | 0.3590 | 0.6025 | 0.4782 |

**Table 1: Results for the ANOVAs – Left hemisphere**. P-values for the 2-way and 3-way ANOVAs on the adjusted-$R^2$ correlation factor computed between VA and the early-stage auditory areas A1, R, L1, L2, L3, L4, M1, M2, M3 and M4. The p-values lower than 0.05 are highlighted in green. The study labelled "a" correspond to Grisendi et al., 2019a, while the study "b" correspond to Grisendi et al., 2019b.

| RH | Study | A1 | R | L1 | L2 | L3 | L4 | M1 | M2 | M3 | M4 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Main effect of voc** | a | 0.0995 | 0.1170 | 0.2241 | 0.0608 | 0.1231 | 0.4936 | 0.0003 | 0.0593 | 0.0815 | 0.3686 |
| | b | 0.0007 | 0.0018 | 0.0036 | 0.0026 | 0.0129 | 0.0205 | 0.4627 | 0.0340 | 0.0255 | 0.0254 |
| **Main effect of val** | a | 0.6707 | 0.9906 | 0.4342 | 0.6301 | 0.5796 | 0.9227 | 0.0069 | 0.5771 | 0.5000 | 0.9533 |
| | b | 0.2334 | 0.4745 | 0.4991 | 0.8002 | 0.9455 | 0.7043 | 0.9962 | 0.2185 | 0.4500 | 0.8361 |
| **Main effect of lat** | a | | | | | | | | | | |
| | b | 0.0377 | 0.0568 | 0.2676 | 0.1019 | 0.2738 | 0.4376 | 0.5619 | 0.0618 | 0.1359 | 0.5667 |
| **Inter voc x val** | a | 0.5743 | 0.8753 | 0.4598 | 0.5219 | 0.7628 | 0.8621 | 0.1365 | 0.3486 | 0.7496 | 0.5493 |
| | b | 0.1406 | 0.3203 | 0.2874 | 0.5765 | 0.8876 | 0.8116 | 0.6747 | 0.4086 | 0.8554 | 0.6171 |
| **Inter voc x lat** | a | | | | | | | | | | |
| | b | 0.0354 | 0.0511 | 0.2156 | 0.0873 | 0.3287 | 0.5134 | 0.2791 | 0.0522 | 0.0121 | 0.0455 |
| **Inter val x lat** | a | | | | | | | | | | |
| | b | 0.2672 | 0.4524 | 0.2999 | 0.6134 | 0.6743 | 0.3785 | 0.7898 | 0.2518 | 0.1871 | 0.4370 |
| **Inter voc x val x lat** | a | | | | | | | | | | |
| | b | 0.0687 | 0.1952 | 0.3542 | 0.4321 | 0.6584 | 0.5364 | 0.5012 | 0.3388 | 0.3845 | 0.3762 |

**Table 2: Results for the ANOVAs – Right hemisphere.** P-values for the 2-way and 3-way ANOVAs on the adjusted-$R^2$ correlation factor computed between VA and the early-stage auditory areas A1, R, L1, L2, L3, L4, M1, M2, M3 and M4. The p-values lower than 0.05 are highlighted in green. The study labelled "a" correspond to Grisendi et al., 2019a, while the study "b" correspond to Grisendi et al., 2019b.

### *Study 2: Processing of emotions and lateralization within AC*

In the second study (Grisendi et al., 2019b), the heatmaps (Figure 3-8) revealed that all our ROIs were globally correlated in all conditions. More specifically the correlations between the different regions were stronger in the left hemisphere than in the right hemisphere. The primary auditory areas (A1 and R) were strongly functionally correlated with the lateral belt areas (L1, L2 and L3 especially), as well as with VA but only for the human vocalizations. The correlations between M4 and all other regions were weaker than between the other regions, and this effect was greater for the vocalizations compared to the non-vocalizations as well as for the left, and for centered presentations compared to the right presentations.

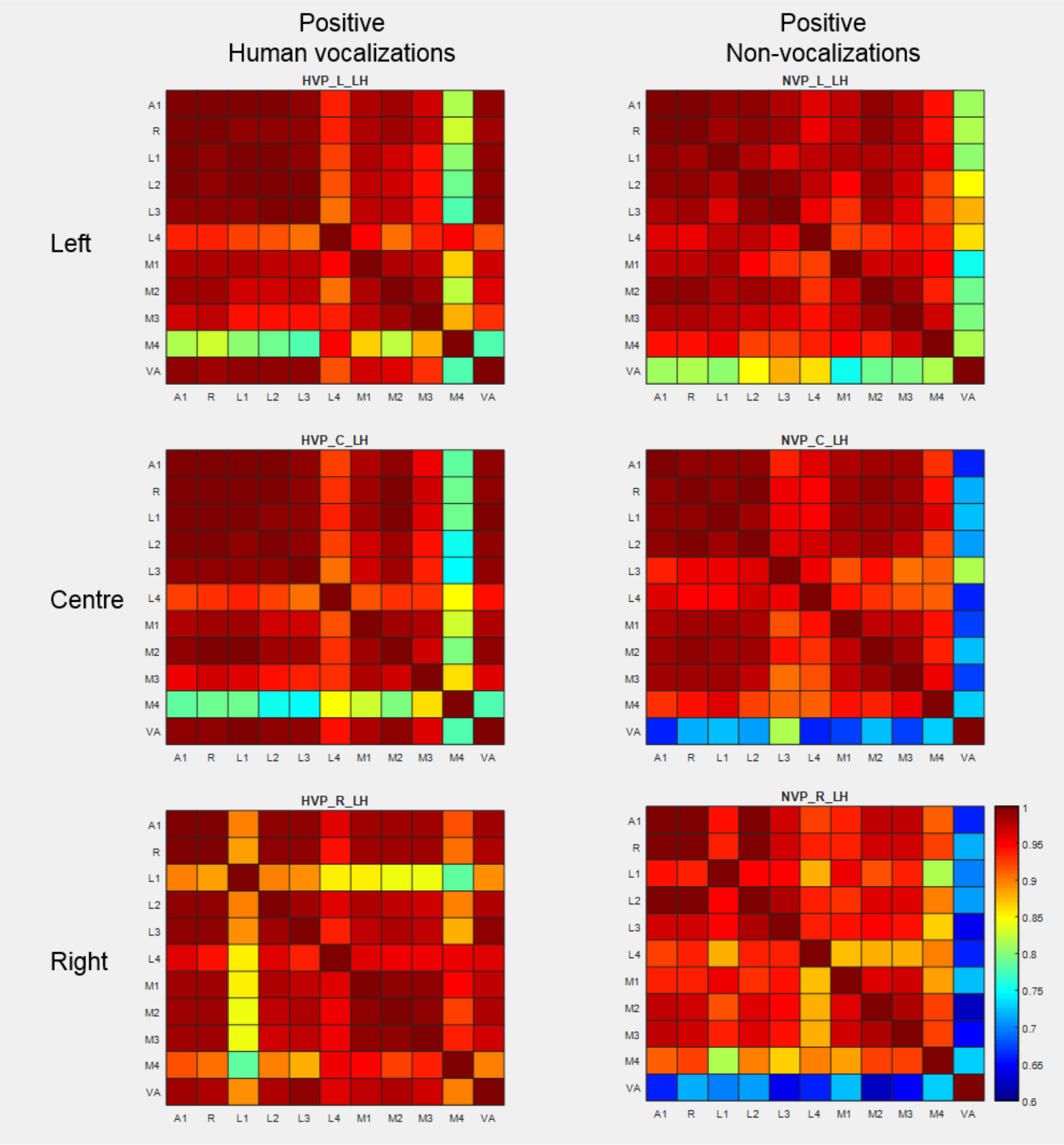

**Figure 3: Heatmaps study 2 left hemisphere positive stimuli**. Heatmaps for the data of the study 2 representing the correlations (adjusted $R^2$) between the BOLD signals of early-stage auditory areas A1, R, L1, L2, L3, L4, M1, M2, M3 and M4 and VA during the presentation of positive human vocalizations (left column) and positive non-vocalizations (right column) with left, centre or right lateralization (top, middle and bottom rows, respectively). The color code represents the strength of the correlation (between [0.6 – 1]).

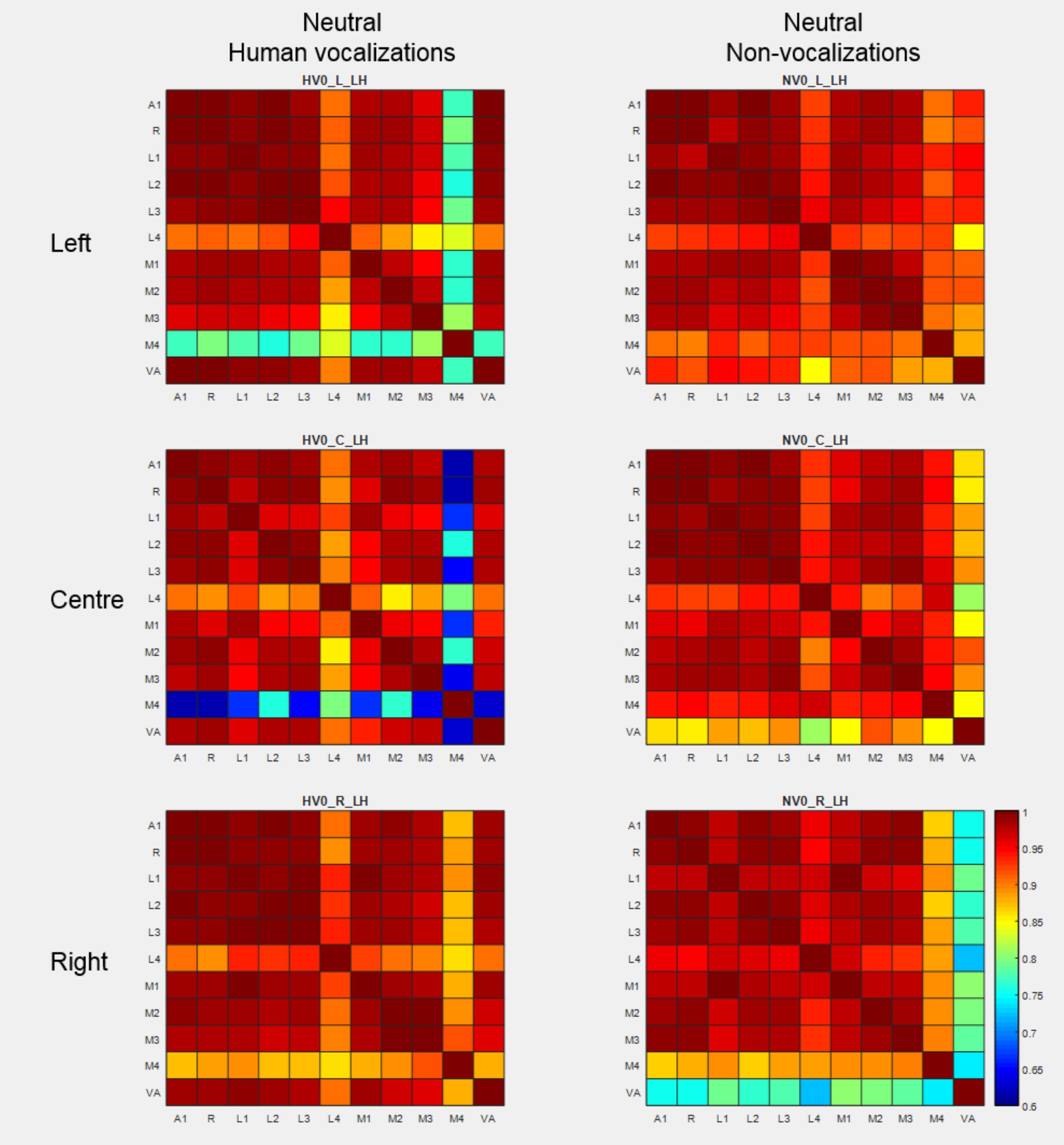

**Figure 4: Heatmaps study 2 left hemisphere neutral stimuli**. Heatmaps for the data of the study 2 representing the correlations (adjusted $R^2$) between the BOLD signals of early-stage auditory areas A1, R, L1, L2, L3, L4, M1, M2, M3 and M4 and VA during the presentation of neutral human vocalizations (left column) and neutral non-vocalizations (right column) with left, centre or right lateralization (top, middle and bottom rows, respectively). The color code represents the strength of the correlation (between [0.6 – 1]).

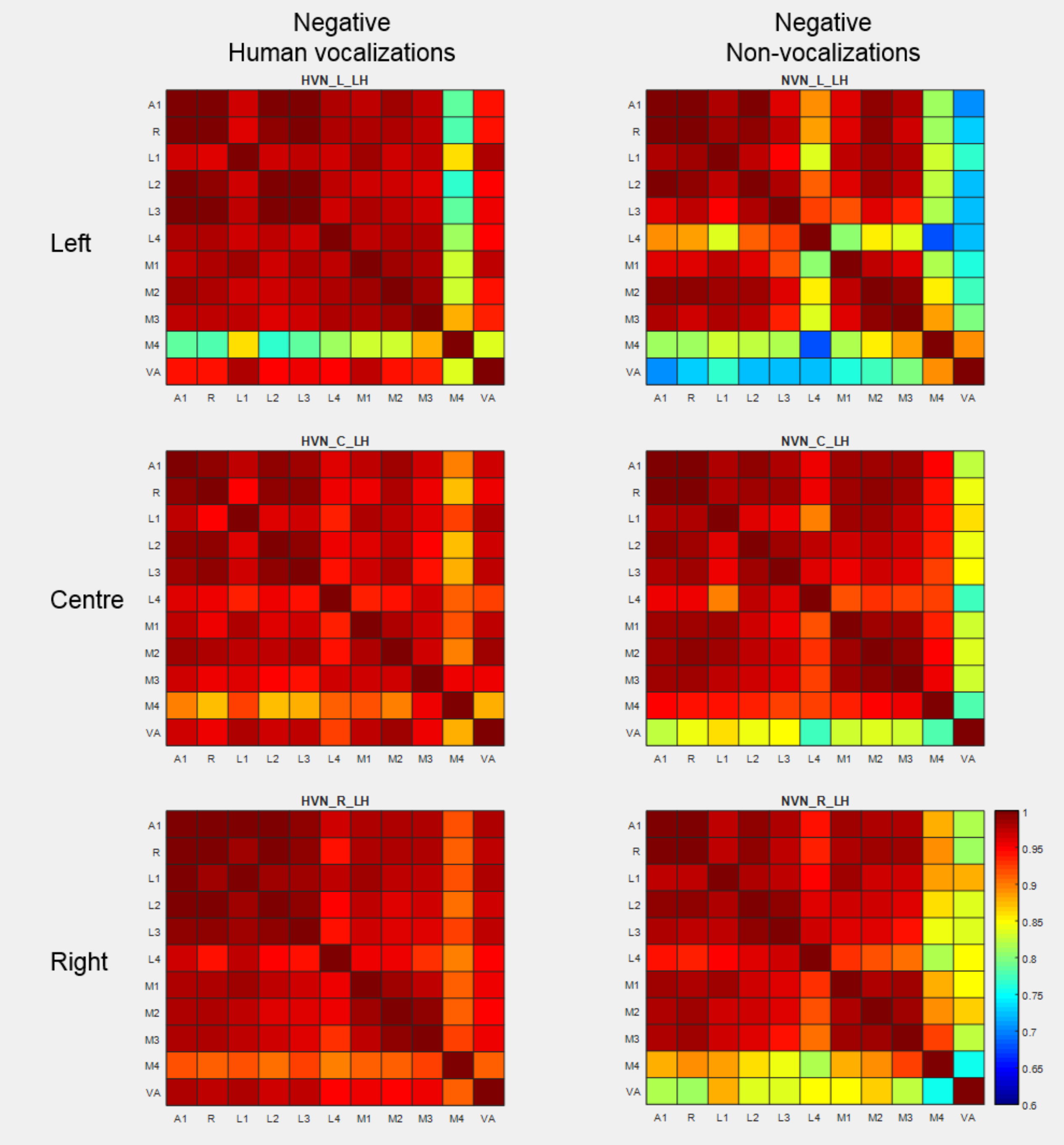

**Figure 5: Heatmaps study 2 left hemisphere negative stimuli**. Heatmaps for the data of the study 2 representing the correlations (adjusted $R^2$) between the BOLD signals of early-stage auditory areas A1, R, L1, L2, L3, L4, M1, M2, M3 and M4 and VA during the presentation of negative human vocalizations (left column) and negative non-vocalizations (right column) with left, centre or right lateralization (top, middle and bottom rows, respectively). The color code represents the strength of the correlation (between [0.6 – 1]).

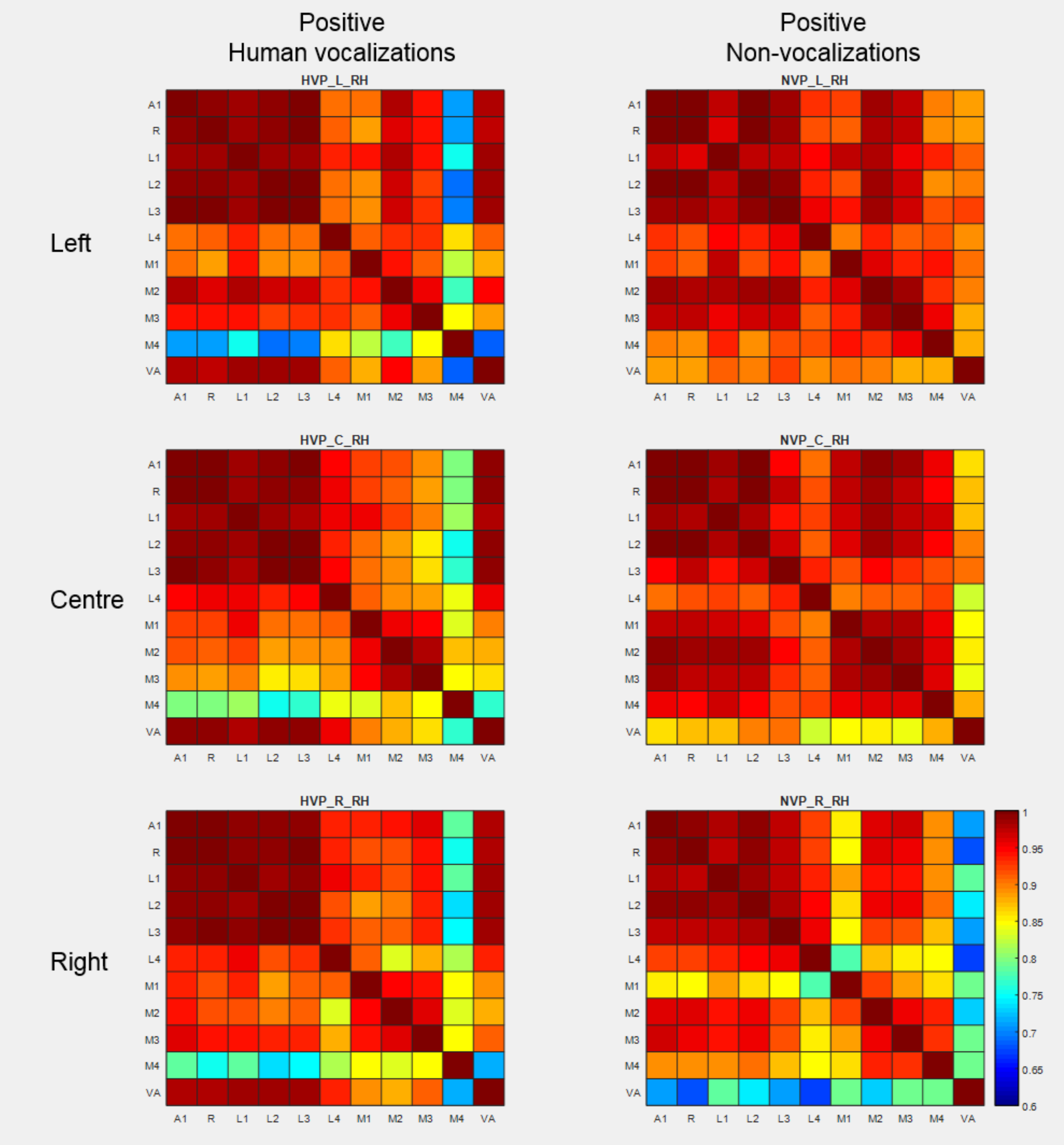

**Figure 6: Heatmaps study 2 right hemisphere positive stimuli**. Heatmaps for the data of the study 2 representing the correlations (adjusted $R^2$) between the BOLD signals of early-stage auditory areas A1, R, L1, L2, L3, L4, M1, M2, M3 and M4 and VA during the presentation of positive human vocalizations (left column) and positive non-vocalizations (right column) with left, centre or right lateralization (top, middle and bottom rows, respectively). The color code represents the strength of the correlation (between [0.6 – 1]).

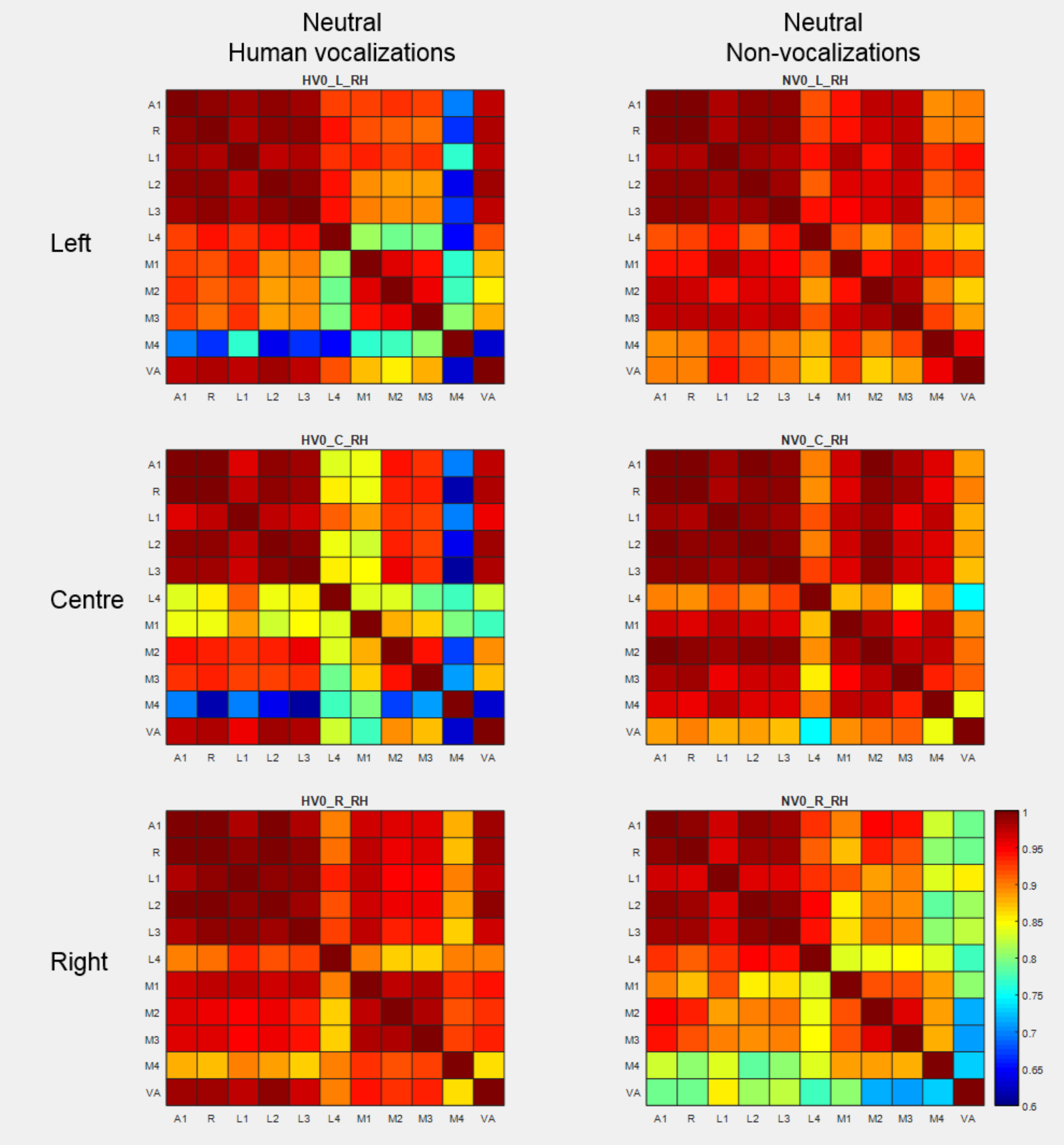

**Figure 7: Heatmpas study 2 right hemisphere neutral stimuli**. Heatmaps for the data of the study 2 representing the correlations (adjusted $R^2$) between the BOLD signals of early-stage auditory areas A1, R, L1, L2, L3, L4, M1, M2, M3 and M4 and VA during the presentation of neutral human vocalizations (left column) and neutral non-vocalizations (right column) with left, centre or right lateralization (top, middle and bottom rows, respectively). The color code represents the strength of the correlation (between [0.6 – 1]).

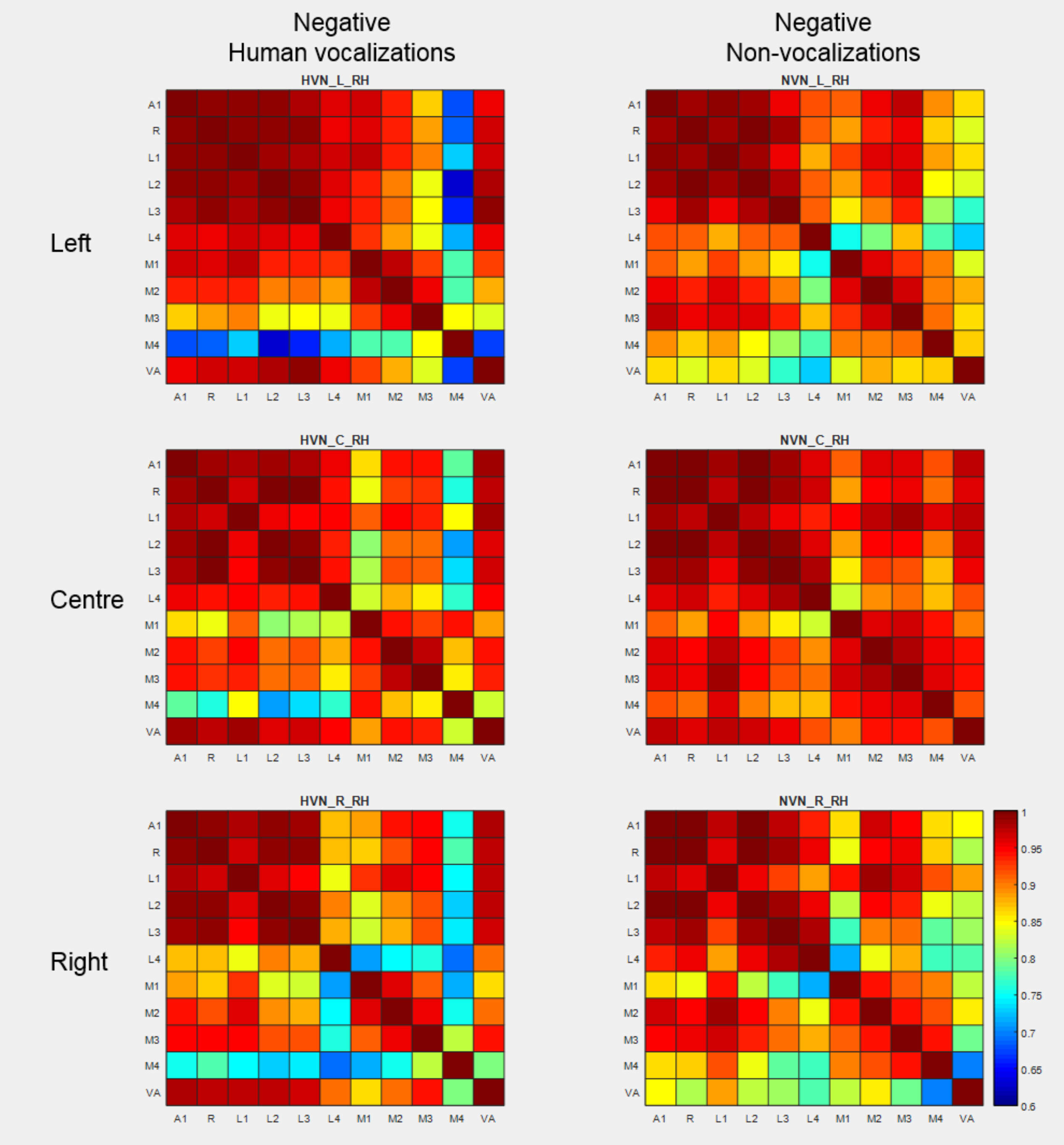

**Figure 8: Heatmaps study 2 right hemisphere negative stimuli**. Heatmaps for the data of the study 2 representing the correlations (adjusted $R^2$) between the BOLD signals of early-stage auditory areas A1, R, L1, L2, L3, L4, M1, M2, M3 and M4 and VA during the presentation of negative human vocalizations (left column) and negative non-vocalizations (right column) with left, centre or right lateralization (top, middle and bottom rows, respectively). The color code represents the strength of the correlation (between [0.6 – 1]).

The 3-way ANOVA: Vocalization (human vocalizations, non-vocalizations) x Valence (negative, neutral, positive) x Lateralization (left, centre, right) revealed a main effect of vocalization bilaterally between VA and all the ROIs. This effect was stronger between VA and, A1 and R, than with the lateral areas (L1, L2, L3 and L4). This effect was due to stronger correlations between VA and early-stage auditory areas for human vocalizations compared to non-vocalizations (Figure 9 and 10). This bias was present in both hemisphere, for all valences and lateralizations (Figure 3-8). A main effect of lateralization was found between VA and A1 in the right hemisphere, as well as an interaction Vocalization x Lateralization between the same regions. The heatmaps revealed that the main effect of lateralization was driven mainly by the non-vocalizations and not by the human vocalizations, that might explain the interaction Vocalization x Lateralization. For the positive and neutral non-vocalizations, the strength of the correlation between VA and A1 was lower for ipsilateral presentations than for centre, and bigger for contralateral presentations (Figure 6 and 7). For the negative non-vocalizations, the correlation was stronger for the centre stimuli compared to the lateralized ones (Figure 8).

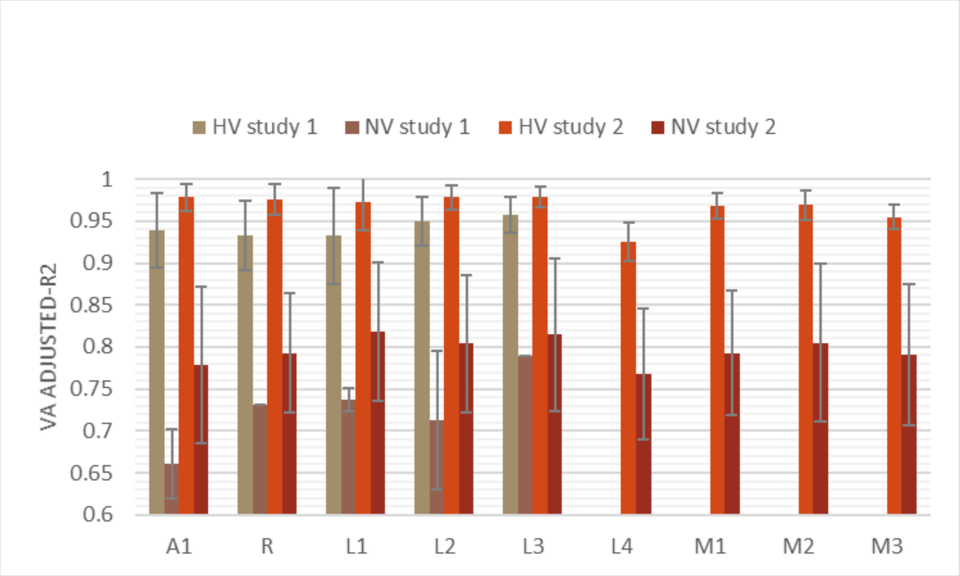

**Figure 9: Significant results for the main effect of Vocalization – Left hemisphere.** Bar plot representing the mean adjusted-R$^2$ between VA and early-stage auditory areas, for the correlations with a significant p-value for the main effect of vocalizations (2-way and 3-way ANOVAs). The sound categories are grouped in human-vocalizations (HV; positive, negative and neutral) or non-vocalizations (NV; positive, negative and neutral). The error bars represent the standard deviation across subjects.

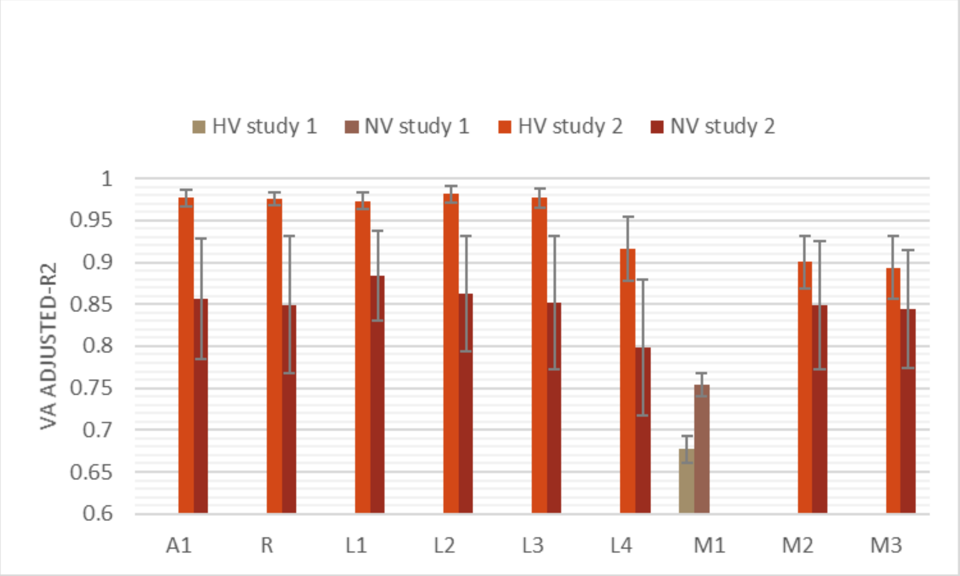

**Figure 10: Significant results for the main effect of Vocalization – Right hemisphere.** Bar plot representing the mean adjusted-$R^2$ between VA and early-stage auditory areas, for the correlations with a significant p-value for the main effect of vocalizations (2-way and 3-way ANOVAs). The sound categories are grouped in human-vocalizations (HV; positive, negative and neutral) or non-vocalizations (NV; positive, negative and neutral). The error bars represent the standard deviation across subjects.

*Comparison study 1 vs. study 2*

Comparing both studies, we observed stronger correlations in the context of a spatial paradigm compared to a non-lateralized experiment. This increased strength was consistent through every type of sounds and valence. The statistical analyses revealed that the same set of areas exhibited a main effect of vocalization in both studies (e.g., A1, R, L1, L2 and L3 in the left hemisphere), but that additional regions are engaged with lateralized sounds.

## Discussion

Our results demonstrated that the correlation of the BOLD signal between VA and individual early-stage auditory areas varied across condition. With human vocalizations stimuli, VA was strongly correlated with regions of PAC and the belt areas, which was not the case for the non-vocalization sounds. Our correlations were in line with functional connectivity results showing an increased connectivity between the regions involved in voice processing for vocal sounds compared to non-vocal sounds (Aglieri et al., 2018). They also described stronger coupling within the left hemisphere than the right hemisphere, as we also reported when using lateralized sounds. However, this left hemispheric dominance in case of lateralized sounds compared with the right hemispheric dominance in case of non-spatial auditory stimuli was a dissimilarity between both studies. The other important point to keep in mind is the big difference in term of strength of correlations. The correlations between our ROIs in the study 2 (with spatial modulation) were stronger than the correlations for the same ROIs in the study 1. This difference could not be only due to the spatial context, as a number of characteristics were different between both studies. Indeed, in the study 2, the number of subject was increased (13 *vs.* 10), the scanner coil has a larger number of channels (32 *vs.* 8), and the VA was defined by mean of a voice localizer. The comparison between the results of study 1 and 2 were descriptive and further investigations would be necessary to infer the impact of using lateralized sounds on the functional correlations between VA and early-stage auditory areas.

The connections between VA and the lateral belt areas were already demonstrated in diffusion spectrum imaging or functional connectivity studies (Cammoun et al., 2015; Pernet et al., 2015). The hierarchical organization of the auditory cortex was established by various studies, where PAC was mainly connected to the surrounding belt areas, which were in turn connected to the parabelt areas (Cammoun et al., 2015; Hackett and Kaas, 2004; LeDoux, 2000). However, their findings limited the connections of VA to the lateral belt areas and not to PAC. The strong functional correlations between VA and PAC may thus be the result of two-step connection between these two regions. Only one previous study described the functional connectivity between PAC with other auditory areas in an emotional context (Koelsch et al., 2018). Using a paradigm with emotional music, their results highlighted a functional connectivity network between regions of auditory association cortex with extra-auditory regions (e.g., limbic, somatosensory, visual, attentional), in contrast with a second network consisting of intra-auditory regions. However, the low spatial resolution (3mm isotropic) used in their study did not allow them to draw a more precise model in the auditory cortex. Another study proposed a dissociation of the connectivity of the STS within the dual-stream model (Erickson et al., 2017). Their findings demonstrated that the anterior regions of the STS were more connected with the ventral stream, while the posterior areas were more connected with the dorsal auditory pathway.

Previous studies suggested a model of dual input to VA, from the lateral belt areas and from the amygdala (Schirmer and Gunter, 2017; Grisendi et al., 2019a), this could not be entirely confirmed in the present study as the amygdala was not part of our analysis. In Grisendi et al., (2019b), the amygdala was not sensible to the spatial cues used. An interesting follow-up study would integrate extra-auditory regions such as amygdala or frontal areas in the analysis. Moreover, in the present study we only investigated intra-hemispheric correlations, as our hypotheses were based on the processing pathways located in the same hemisphere, thus to infer any inter-hemispheric effects, future experiments could investigate evidence for interhemispheric connections within the auditory cortex, as well as between the auditory cortex and other areas of the auditory processing stream (Budinger and Heil, 2006; Kaas and Hackett, 2005).

## Conclusions

The present exploratory study highlighted a modulation of the coupling of VA and early-stage auditory areas in function of the category of sound. Human vocalizations induce stronger correlations between VA, PAC and lateral belt areas. Thus, the functional correlations between the auditory areas was not restricted to correlations between PAC and belt areas or between belt areas and parabelt areas, but comprised also strong functional correlations between PAC and VA for the specific processing of human vocalizations.

## References

Aeschlimann, M., Knebel, J.-F., Murray, M.M., Clarke, S., 2008. Emotional Pre-eminence of Human Vocalizations. Brain Topogr. 20, 239–248. https://doi.org/10.1007/s10548-008-0051-8

Aglieri, V., Chaminade, T., Takerkart, S., Belin, P., 2018. Functional connectivity within the voice perception network and its behavioural relevance. NeuroImage 183, 356–365. https://doi.org/10.1016/j.neuroimage.2018.08.011

Arnal, L.H., Flinker, A., Kleinschmidt, A., Giraud, A.-L., Poeppel, D., 2015. Human Screams Occupy a Privileged Niche in the Communication Soundscape. Curr. Biol. 25, 2051–2056. https://doi.org/10.1016/j.cub.2015.06.043

Beaucousin, V., Lacheret, A., Turbelin, M.-R., Morel, M., Mazoyer, B., Tzourio-Mazoyer, N., 2007. FMRI Study of Emotional Speech Comprehension. Cereb. Cortex 17, 339–352. https://doi.org/10.1093/cercor/bhj151

Belin, P., Zatorre, R.J., Ahad, P., 2002. Human temporal-lobe response to vocal sounds. Cogn. Brain Res. 13, 17–26. https://doi.org/10.1016/S0926-6410(01)00084-2

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312. https://doi.org/10.1038/35002078

Bestelmeyer, P.E.G., Kotz, S.A., Belin, P., 2017. Effects of emotional valence and arousal on the voice perception network. Soc. Cogn. Affect. Neurosci. 12, 1351–1358. https://doi.org/10.1093/scan/nsx059

Budinger, E., Heil, P., 2006. Anatomy of the auditory cortex. pp. 91–113.

Cammoun, L., Thiran, J.P., Griffa, A., Meuli, R., Hagmann, P., Clarke, S., 2015. Intrahemispheric cortico-cortical connections of the human auditory cortex. Brain Struct. Funct. 220, 3537–3553. https://doi.org/10.1007/s00429-014-0872-z

Chiry, O., Tardif, E., Magistretti, P.J., Clarke, S., 2003. Patterns of calcium-binding proteins support parallel and hierarchical organization of human auditory areas. Eur. J. Neurosci. 17, 397–410. https://doi.org/10.1046/j.1460-9568.2003.02430.x

Clarke, S., Morosan, P., 2012. Architecture, connectivity and transmitter receptors of human auditory cortex, in: Poeppel, D., Overath, T., Popper, A. N., and Fay, R.R. (Eds). (2012). Human Auditory Cortex. pp. 11–38.

Clarke, S., Rivier, F., 1998. Compartments within human primary auditory cortex: evidence from cytochrome oxidase and acetylcholinesterase staining. Eur. J. Neurosci. 10, 741–745. https://doi.org/10.1046/j.1460-9568.1998.00043.x

Da Costa, S., Bourquin, N.M.-P., Knebel, J.-F., Saenz, M., Zwaag, W. van der, Clarke, S., 2015. Representation of Sound Objects within Early-Stage Auditory Areas: A Repetition Effect Study Using 7T fMRI. PLOS ONE 10, e0124072. https://doi.org/10.1371/journal.pone.0124072

Da Costa, S., Zwaag, W. van der, Marques, J.P., Frackowiak, R.S.J., Clarke, S., Saenz, M., 2011. Human Primary Auditory Cortex Follows the Shape of Heschl's Gyrus. J. Neurosci. 31, 14067–14075. https://doi.org/10.1523/JNEUROSCI.2000-11.2011

Da Costa, S., Clarke, S., Crottaz-Herbette, S., 2018. Keeping track of sound objects in space: The contribution of early-stage auditory areas. Hear. Res., International Conference on Auditory Cortex 2017 366, 17–31. https://doi.org/10.1016/j.heares.2018.03.027

Erickson, L.C., Rauschecker, J.P., Turkeltaub, P.E., 2017. Meta-analytic connectivity modeling of the human superior temporal sulcus. Brain Struct. Funct. 222, 267–285. https://doi.org/10.1007/s00429-016-1215-z

Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., Grodd, W., Wildgruber, D., 2006. Effects of prosodic emotional intensity on activation of associative auditory cortex: NeuroReport 17, 249–253. https://doi.org/10.1097/01.wnr.0000199466.32036.5d

Ethofer, T., Van De Ville, D., Scherer, K., Vuilleumier, P., 2009. Decoding of Emotional Information in Voice-Sensitive Cortices. Curr. Biol. 19, 1028–1033. https://doi.org/10.1016/j.cub.2009.04.054

Frühholz, S., Trost, W., Kotz, S.A., 2016. The sound of emotions—Towards a unifying neural network perspective of affective sound processing. Neurosci. Biobehav. Rev. 68, 96–110. https://doi.org/10.1016/j.neubiorev.2016.05.002

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., Vuilleumier, P., 2005. The voices of wrath: brain responses to angry prosody in meaningless speech. Nat. Neurosci. 8, 145–146. https://doi.org/10.1038/nn1392

Grisendi T, Reynaud O, Clarke S, Da Costa S 2019a. Processing pathways for emotional vocalizations. Brain Struct Funct, In revision.

Grisendi T, Clarke S, Da Costa S 2019b. Emotional value of the auditory space. In preparation.

Hackett, T.A., Kaas, J.H., 2004. Auditory Cortex in Primates: Functional Subdivisions and Processing Streams, in: The Cognitive Neurosciences, 3rd Ed. MIT Press, Cambridge, MA, US, pp. 215–232.

Hackett, T.A., Preuss, T.M., Kaas, J.H., 2001. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. J. Comp. Neurol. 441, 197–222. https://doi.org/10.1002/cne.1407

Kaas, J.H., Hackett, T.A., 2005. Subdivisions and connections of the auditory cortex in primates: A working model. pp. 7–26.

Koelsch, S., Skouras, S., Lohmann, G., 2018. The auditory cortex hosts network nodes influential for emotion processing: An fMRI study on music-evoked fear and joy. PLoS ONE 13. https://doi.org/10.1371/journal.pone.0190057

Lavan, N., Rankin, G., Lorking, N., Scott, S., McGettigan, C., 2017. Neural correlates of the affective properties of spontaneous and volitional laughter types. Neuropsychologia 95, 30–39. https://doi.org/10.1016/j.neuropsychologia.2016.12.012

LeDoux, J.E., 2000. Emotion Circuits in the Brain. Annu. Rev. Neurosci. 23, 155–184. https://doi.org/10.1146/annurev.neuro.23.1.155

Leitman, D.I., Wolf, D.H., Ragland, J.D., Laukka, P., Loughead, J., Valdez, J.N., Javitt, D.C., Turetsky, B.I., Gur, R.C., 2010. "It's Not What You Say, But How You Say it": A Reciprocal Temporo-frontal Network for Affective Prosody. Front. Hum. Neurosci. 4. https://doi.org/10.3389/fnhum.2010.00019

Obleser, J., Zimmermann, J., Van Meter, J., Rauschecker, J.P., 2007. Multiple Stages of Auditory Speech Perception Reflected in Event-Related fMRI. Cereb. Cortex 17, 2251–2257. https://doi.org/10.1093/cercor/bhl133

Pernet, C.R., McAleer, P., Latinus, M., Gorgolewski, K.J., Charest, I., Bestelmeyer, P.E.G., Watson, R.H., Fleming, D., Crabbe, F., Valdes-Sosa, M., Belin, P., 2015. The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. NeuroImage 119, 164–174. https://doi.org/10.1016/j.neuroimage.2015.06.050

Rivier, F., Clarke, S., 1997. Cytochrome Oxidase, Acetylcholinesterase, and NADPH-Diaphorase Staining in Human Supratemporal and Insular Cortex: Evidence for Multiple Auditory Areas. NeuroImage 6, 288–304. https://doi.org/10.1006/nimg.1997.0304

Schirmer, A., Gunter, T.C., 2017. Temporal signatures of processing voiceness and emotion in sound. Soc. Cogn. Affect. Neurosci. 12, 902–909. https://doi.org/10.1093/scan/nsx020

Schirmer, A., Kotz, S.A., 2006. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. Trends Cogn. Sci. 10, 24–30. https://doi.org/10.1016/j.tics.2005.11.009

van der Zwaag, W., Gentile, G., Gruetter, R., Spierer, L., Clarke, S., 2011. Where sound position influences sound object representations: A 7-T fMRI study. NeuroImage 54, 1803–1811. https://doi.org/10.1016/j.neuroimage.2010.10.032

Viceic, D., Fornari, E., Thiran, J.-P., Maeder, P.P., Meuli, R., Adriani, M., Clarke, S., 2006. Human auditory belt areas specialized in sound recognition: a functional magnetic resonance imaging study. Neuroreport 17, 1659–1662. https://doi.org/10.1097/01.wnr.0000239962.75943.dd

Wallace, M.N., Johnston, P.W., Palmer, A.R., 2002. Histochemical identification of cortical areas in the auditory region of the human brain. Exp. Brain Res. 143, 499–508. https://doi.org/10.1007/s00221-002-1014-z

Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., Ackermann, H., 2004. Distinct Frontal Regions Subserve Evaluation of Linguistic and Emotional Aspects of Speech Intonation. Cereb. Cortex 14, 1384–1389. https://doi.org/10.1093/cercor/bhh099