

# Computational neuroimaging of real-life listening

## Citation for published version (APA):

De Angelis, V. (2019). *Computational neuroimaging of real-life listening*. Maastricht University. <https://doi.org/10.26481/dis.20190410vda>

## Document status and date:

Published: 01/01/2019

## DOI:

[10.26481/dis.20190410vda](https://doi.org/10.26481/dis.20190410vda)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

## Chapter 5

### General discussion



## Main findings and open questions

The main goal of the present thesis was to investigate how real-life sounds and sound mixtures are represented and processed in the human auditory cortex. In particular, the computational mechanisms to derive higher-level sound and auditory scene representations from their lower level acoustic representations were investigated through a combination of modeling and functional MRI data analysis.

In the study reported in Chapter 2, the combination of a pitch processing model with fMRI revealed an auditory cortical region where an explicit representation of pitch height and salience of complex sounds are calculated. This region is located along the lateral HG/middle STG and its location is consistent with reports of previous studies investigating pitch processing with synthetic sounds. Moreover, the integration of both pitch height and salience to a multi-resolution spectro-temporal stimulus representation model provided a higher performance in decoding complex sounds from the fMRI activation patterns, especially from that specific region. These findings support the functional role of a specific site of the human auditory cortex in pitch processing; however the issue concerning whether a temporal or a spectral mechanism drives the corresponding neuronal computations remains an open question. The temporal resolution of the fMRI signal (Logothetis et al., 2001) represents an intrinsic limitation in investigating the faster temporal dynamics of neuronal phase-locking and its degradation within the auditory system observed in animals (from 2-4 kHz in the auditory nerve to 100-200 Hz in the auditory cortex, see Oxenham, 2013). Such loss of temporal information further suggests that pitch might be processed - at least to a certain extent - at an earlier (sub-cortical) stage of the ascending auditory pathway (Oxenham, 2013; Shamma and Klein, 2000). Functional measurements at high spatio-temporal resolution and including both subcortical and cortical structures as well as modeling of neural population temporal dynamics would be needed to addressing this open issue.

Chapter 3 investigated how higher-level sound representations (sound categories) are encoded in the fMRI responses. The representation of sound categories was modeled as a linear combination of relevant low-level acoustic (frequency, spectro-temporal modulations) and perceptual (pitch) sound attributes, weighted by their discriminative power learned in a prior classification task. Overall, the fMRI responses of the HG/PT/PP showed a preferential encoding of the acoustic information of the input stimuli, whereas sound cate-

gories were identified with higher accuracy from the fMRI activity of secondary areas consisting of the STG/STS. This does not directly imply the complete loss of acoustical information in these last regions, as the fine-grained low-level representation may still be relevant to identify distinct sounds within the same class.

Taken together, these observations suggest that alternative models enhancing the representational distance between sounds of different categories and preserving the within-class acoustic information may better explain the neural transformations occurring within the STG/STS regions. The results of this study are limited by the absence of behavioral tasks specifically related to the presented sound categories. Thus, the approach of testing different models, as described here, may be further extended to analyze fMRI data collected while subjects perform specific cognitive tasks (e.g. sound categorization in pre-defined categories, within-category sound identification). This would allow investigating whether and how the behavioral requirements affect the representation of sounds in primary and non-primary auditory areas.

The aim of Chapter 4 was to investigate - in the context of a biologically-inspired computational model of ASA - the role of low-level acoustic/perceptual sound attributes and top-down grouping principles in processing real-life sound mixtures. The model included acoustic and perceptual features that in Chapter 2 and Chapter 3 resulted encoded in the fMRI signal of the auditory cortex (frequency, spectro-temporal modulations, pitch). An interesting result in this chapter was that highest performances of the model were obtained by including prior knowledge at the level of sound category as a mechanism that modulates the sensitivity of neural responses to the acoustic feature.

This modeling work provides a framework for future neuroimaging studies. For instance, fMRI responses to complex scenes could be collected while subjects perform specific cognitive tasks (i.e. attend one specific source in a mixture). Similar to previous electrophysiological studies (e.g. Mesgarani and Chang, 2012), the measured fMRI activity could be then used to reconstruct the attended stream and the reconstruction results compared to the outputs of the model to assess the correspondence between model and brain representation of foreground vs background sounds. Furthermore, the comparison of the responses measured in different auditory areas with the output of the model at each computational stage would allow investigating where, in the human auditory cortex, the different stages of processing complex scenes may take place.

## **Future directions and methodological developments**

In Chapter 2 and Chapter 3 the cortical processing of low-level sensory features and the relation with higher-level semantic representations of natural sounds were investigated through the combination of linearized encoding and decoding models (see Naselaris et al., 2011 for more details). The methods implemented in these two chapters extended those of previous work that aimed at characterizing the auditory cortex in terms of processing of the acoustic features alone (Moerel et al., 2012; Santoro et al., 2014, 2017). This approach is interesting as it provides a rigorous description of the neural computations that may underly the observations reported in previous neuroimaging studies adopting a more conventional contrast-based approach (i.e. Belin et al., 2000; Binder, 2000; Griffiths and Hall, 2012; Leaver and Rauschecker, 2010; Patterson et al., 2002; Penagos et al., 2004).

The results obtained in this thesis support the suitability of the adopted procedures to investigate multiple processing mechanisms within the human auditory cortex. However, the information explained from the measured cortical responses depends on both the tested hypotheses (i.e. the computational models) and on the fMRI experimental design. On the modeling side, future developments could be oriented towards the identification of higher-dimensional feature spaces combining multiple levels of information (acoustic, perceptual, semantic) in complement to the features considered in the present thesis. This could be done, e.g. using the approach of deep neural networks (DNN). Recently, the popularity of DNNs has increased due to the high capability in replicating human behavior in specific auditory or visual processing tasks (Güçlü and van Gerven, 2015; Kell et al., 2018; Yamins et al., 2014). Although the computational structure and the requirements of extended dataset to train the models may not fit the biological constraints of the human brain (Kell et al., 2018), DNNs may be useful in discovering and specifying higher level representations and generate hypotheses that can be tested with neuroimaging. Moreover, as the sensitivity of neural populations is modulated by the execution of cognitive tasks (Atiani et al., 2009; Fritz et al., 2003), it would be important to consider how these feature spaces and representations are modulated by the specific behavioral context and how analytical approaches (e.g. based on fMRI-encoding) can be extended to consider such modulations. Finally, the computational modeling framework presented in Chapter 4 can be used to extend current analyses of fMRI (and combined fMRI-EEG) data, which will enable investigating the

contribution of bottom-up and top-down computational mechanisms underlying auditory scene analysis.

In summary, the studies presented in this thesis show that the combination of biologically-inspired computational modeling and high-field fMRI is a powerful instrument to unravel the complex processing mechanisms of the human brain. Furthermore, the methods described can be extended towards new directions of scientific investigations. Due to the lack of direct measures of the neural computations, the development of computational models capable to describe into details the complexity of the human brain remains challenging. The integration of different measurement methods ensuring both high spatial and high temporal resolution as well as the integration of models at different levels, from biophysical to functional models, is a necessary pre-requisite to achieve this goal.

---

## References

- Atiani S, Elhilali M, David SV, Fritz JB, and Shamma SA (2009). Task Difficulty and Performance Induce Diverse Adaptive Patterns in Gain and Shape of Primary Auditory Cortical Receptive Fields. *Neuron* 61(3), 467–480. DOI: 10.1016/j.neuron.2008.12.027.
- Belin P, Zatorre R J, Lafaille P, Ahad P, and Pike B (2000). Voice-selective areas in human auditory cortex. *Nature* 403(6767), 309–312. DOI: 10.1038/35002078.
- Binder JR (2000). Human Temporal Lobe Activation by Speech and Nonspeech Sounds. *Cereb Cortex* 10(5), 512–528. DOI: 10.1093/cercor/10.5.512.
- Fritz J, Shamma SA, Elhilali M, and Klein D (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* 6, 1216.
- Griffiths TD and Hall DA (2012). Mapping pitch representation in neural ensembles with fMRI. *J Neurosci* 32(39), 13343–7. DOI: 10.1523/jneurosci.3813-12.2012.
- Güçlü U and van Gerven MAJ (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *J Neurosci* 35(27), 10005–10014. DOI: 10.1523/JNEUROSCI.5023-14.2015.
- Kell AJE, Yamins DLK, Shook EN, Norman-Haignere SV, and McDermott JH (2018). A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy. *Neuron* 98(3), 630–644.e16. DOI: 10.1016/j.neuron.2018.03.044.
- Leaver AM and Rauschecker JP (2010). Cortical Representation of Natural Complex Sounds: Effects of Acoustic Features and Auditory Object Category. *J Neurosci* 30(22), 7604–7612. DOI: 10.1523/JNEUROSCI.0296-10.2010.
- Logothetis NK, Pauls J, Augath M, Trinath T, and Oeltermann A (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412. DOI: 10.1038/35084005.
- Mesgarani N and Chang EF (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485(7397). DOI: 10.1038/nature11020.
- Moerel M, De Martino F, and Formisano E (2012). Processing of Natural Sounds in Human Auditory Cortex: Tonotopy, Spectral Tuning, and Relation to Voice Sensitivity. *J Neurosci* 32(41), 14205–14216. DOI: 10.1523/JNEUROSCI.1388-12.2012.

- Naselaris T, Kay KN, Nishimoto S, and Gallant JL (2011). Encoding and Decoding in fMRI. *NeuroImage* 56(2), 400–410. DOI: 10.1016/j.neuroimage.2010.07.073.
- Oxenham AJ (2013). Revisiting place and temporal theories of pitch. *Acoust Sci Technol* 34(6), 388–396.
- Patterson RD, Uppenkamp S, Johnsrude IS, and Griffiths TD (2002). The Processing of Temporal Pitch and Melody Information in Auditory Cortex. *Neuron* 36(4), 767–776.
- Penagos H, Melcher JR, and Oxenham AJ (2004). A Neural Representation of Pitch Salience in Nonprimary Human Auditory Cortex Revealed with Functional Magnetic Resonance Imaging. *J Neurosci* 24(30), 6810–6815. DOI: 10.1523/JNEUROSCI.0383-04.2004.
- Santoro R, Moerel M, De Martino F, Goebel R, Ugurbil K, Yacoub E, and Formisano E (2014). Encoding of Natural Sounds at Multiple Spectral and Temporal Resolutions in the Human Auditory Cortex. *PLoS Comp Biol* 10(1). DOI: 10.1371/journal.pcbi.1003412.
- Santoro R, Moerel M, De Martino F, Valente G, Ugurbil K, Yacoub E, and Formisano E (2017). Reconstructing the spectrotemporal modulations of real-life sounds from fMRI response patterns. *Proc Natl Acad Sci USA* 114(18), 4799–4804. DOI: 10.1073/pnas.1617622114.
- Shamma SA and Klein D (2000). The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. *J Acoust Soc Am* 107(5), 2631–2644. DOI: 10.1121/1.428649.
- Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, and DiCarlo JJ (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc Natl Acad Sci USA* 111(23), 8619–8624. DOI: 10.1073/pnas.1403112111.