



THE PROCESSING OF TEMPORAL PITCH AND MELODY INFORMATION IN AUDITORY CORTEX - REVISITED

Stefan Uppenkamp*

Emily Bünker

Dominik Langner

Anne zur Horst

Department für Medizinische Physik und Akustik, Fakultät VI
Carl von Ossietzky Universität Oldenburg
26111 Oldenburg, Germany

ABSTRACT

In 2002 Patterson et al. published a highly cited auditory fMRI study on the processing of sequences of musical notes, realized regular interval sounds presented to both ears (Neuron **36**, 767-776, 2002). The main findings were (1) the identification of a pitch sensitive region in lateral Heschl's gyrus (HG), outside primary auditory cortex, in both hemispheres, and (2) the demonstration of a specific effect of pitch changes, as in melodies, in several adjacent areas in Planum polare and superior temporal gyrus, which for most listeners was more prominent in the right hemisphere. One purpose of the current fMRI study was to shed more light on the hemispheric asymmetry in melody processing. The effect of the ear of entry on the asymmetry of the melody-specific activation was investigated for monaural presentation. Simple contrasts between sound and silence revealed the expected crossed pathway of the main projections between ear and auditory cortex. Melody-specific activation however was largely independent from the ear of entry. This is interpreted as evidence for a hierarchy in pitch processing. Regions that are specifically activated by changes of pitch appear to represent the perceptual component of listening to melodies, beyond the purely sensory representation driven by stimulus properties only.

Keywords: *pitch, hemispheric asymmetry, auditory fMRI*

*Corresponding author: stefan.uppenkamp@uol.de.

Copyright: ©2023 Uppenkamp et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1. INTRODUCTION

Changes of pitch are the basis for musical melodies, and also the different languages in the world use voice pitch to a various extent as one parameter carrying the information of the meaning or at least the general mood of a spoken utterance. A useful operational rather than a verbal definition of pitch was suggested by Hartmann [1]: "We say that a sound has a certain pitch if it can be reliably matched by adjusting the frequency of a sine wave of arbitrary amplitude." This definition also supplies the physical unit for measuring pitch, as pitch is the perceptual correlate of the periodicity of a signal, measured in Hertz ($=s^{-1}$). Musical pitch covers the range of about 30 Hz to about 5 or 6 kHz. Below 30 Hz, it is getting increasingly difficult to decide which of two musical notes was lower or higher in pitch [2, 3], and only few musical instruments produce tones with a fundamental frequency of more than 6 kHz.

The pitch of a sinusoid is largely determined by its frequency peak. Therefore, it is obvious that many pitch phenomena can simply be explained by the auditory system effectively acting as a frequency analyzer, by means of the place-frequency mapping in the cochlea and the preservation of this mapping in the ascending auditory pathway up to auditory cortex. However, several pitch phenomena are more a reflection of the temporal regularity of a sound, e.g., a periodic click train or a regular interval sound such as iterated rippled noise (IRN) [4], rather than determined by spectrally resolved peaks. There is a long history of scientific debate about spectral and temporal pitch models, that both have been more or less successful in describing the various pitch phenomena (see, e.g., [5]). One auditory model making use of time interval processing is the auditory image model suggested by Patterson and

colleagues, which extracts periodic events in the sound waveform and stores this for a limited period of time in an auditory image buffer, to serve as a template for pitch processing [6, 7].

Griffiths et al. [8] had employed iterated rippled noise stimuli in an auditory fMRI study to measure the neural activity associated with the temporal structure of sound in the ascending human auditory pathway from the cochlear nucleus to the cortex. The temporal structure included regularities at the millisecond level and pitch sequences at the hundreds-of-milliseconds level. It was shown that the process of recoding temporal patterns into a more stable form begins as early as the cochlear nucleus and continues up to auditory cortex. There was no effect of pitch variance, like in melodies, at a stage before cortex.

A detailed analysis at cortical level of this dataset gathered from nine participants had been published by Patterson et al. in a second, highly cited paper [9]. That study revealed a hierarchy of pitch processing, with a comparatively small localized pitch specific region bilateral in the antero-lateral aspect of Heschl's gyrus (HG), and melody specific activation clearly outside primary auditory regions in superior temporal gyrus and sulcus and in at the pole of the temporal lobes. This melody-specific activation was more pronounced in the right than in the left hemisphere for most listeners. Although the variability across the participants in that study was comparatively large, especially for the melody-specific activation patterns, the general organizing principles for the processing of pitch and melody sequences appear to be largely consistent. The localization of the pitch specific activation in antero-lateral HG is a finding that was replicated in several follow-up studies by this author [10] and also in other studies [11-13]. It appears as if the lateral part of HG is acting as a pitch specific processing site, irrespective of the exact type of stimuli employed. Evidence for this finding also comes from a series of other studies, including MEG [14-16] and electrophysiological recordings from macaque monkeys [17] and from human listeners [18].

The anatomy and physiology of the human brain are characterized by crossed pathways between the two ears and the two hemispheres of the brain, with the main projections from the ears leading to the respective contralateral hemisphere [19]. The symmetry between hemispheres of the cortical activation maps in response to diotic sounds with fixed temporal pitch, as reported in [9], indicates a symmetry in the neural structures for temporal processing along the auditory pathway and in the unimodal sensory areas in auditory cortex. The

extraction of temporal pitch per se appears to be reflected mainly by sensory coding in these structures. In contrast, the reported asymmetry between left and right hemispheres for the melody specific processing even in the case of diotic stimulation [9] suggests a process that is higher up in the hierarchy, involving cognitive processing, which not only reflects stimulus driven neural excitation in the primary sensory areas but also the way we are listening to melodies.

In the current study, 20 years on from the study reported by Patterson et al. [9], fMRI was used to explore the influence of monaural stimulus presentation on the reported hemispheric asymmetry for the processing of sequences of notes with changing pitch. At the same time, this study is also aimed at replicating the previous findings with a modern 3-Tesla MRI-scanner, using a different high-fidelity sound delivery system, and with more participants from a fairly homogenous group of listeners, allowing for a second-level data analysis across the whole group.

2. METHODS

Stimuli. Iterated rippled noise stimuli (IRN) were used throughout the experiments. IRN is created by delaying a copy of random noise and adding it back to the original. The resulting sound has some of the hiss of the original noise, but it also has a pitch corresponding to the inverse of the delay [4]. The pitch strength increases when the delay-and-add process is repeated. When the pitch is less than about 125 Hz (corresponding to a delay of 8 ms) and the stimuli are high-pass filtered at about 500 Hz, the IRN effectively excites all frequency channels in a similar way as random noise, with no resolved spectral peaks internally. The perception of IRN pitch is therefore assumed to be based on extracting time-intervals from the signal rather than spectral peaks [6, 4].

Sound conditions. The different stimulus conditions in this study included melodies and sequences of fixed-pitch notes, with diotic stimulation of both ears, and monaural stimulation just to the left and just to the right ear. Diotic presentation of random noise bursts (no pitch) and silence were included as controls, giving a total of eight stimulus conditions, four of which similar to the original work [8, 9]. The sounds were played as sequences of 32 notes at a rate of four per second. Each note had a duration of 210 ms with 40 ms of silence between successive notes. The pitch range for the melodies was 50 to 110 Hz. The pitch in the 'fixed-pitch' sequences was varied randomly between sequences

to cover the same range as the melodies over the course of the experiment. All stimuli were bandpass filtered between 500 Hz and 4 kHz and presented to the listeners via MR-compatible headphones (Optoacoustics) at a level of approximately 70 dB SPL. All conditions were repeated 32 times in random order throughout the experiment.

Participants. Fourteen normal-hearing listeners volunteered as participants (six males, eight females, all students aged 21.3 ± 1.2 years). None of the listeners had any history of hearing disorders or neurological disorders.

fMRI methods. Functional images were obtained using a T2*-weighted gradient echo planar imaging (EPI) sequence with an echo time of 30 msec and a flip angle of 90° , using a 3-Tesla MRI scanner (Siemens Prisma). Sparse imaging was employed with a repetition time TR of 11 s and a volume acquisition time of 2.5 s, to separate the scanner noise and the experimental sounds in time [20]. Each volume comprised 40 slices (matrix size 68×66 , 3 mm voxel resolution) with a slice thickness of 3 mm and 15% spacing between slices. The acquired volumes were slightly angled away from the eyes and covered the full brain for all listeners. Four runs with a total of 264 functional images were collected from each participant. After completion of the activation data, a high-resolution structural T1-weighted MRI scan was obtained for each participant ($320 \times 320 \times 224$ voxels at resolution of 0.75 mm voxel size).

Data analysis. Anatomical and functional data were analyzed using SPM12 (<http://www.fil.ion.ucl.ac.uk/spm>). The preprocessing of the BOLD images included realignment of subject motion, normalization of individual scans to a standard EPI template, and smoothing with a Gaussian filter of 5 mm full width at half maximum. Statistical analysis of the data was done in two steps. At the first level, a voxel-wise general linear model with the eight sound conditions as well as the movement parameters as additional regressors was set up for each individual listener, to find an estimate for the time-course the stimulus-related BOLD response for each condition. Contrast images for each contrast of interest were calculated. From these, statistical maps may be created, indicating the significance of differences between the experimental conditions, usually denoted as individual activation maps. For the group analysis, contrast images for each effect of interest from all individuals are fed into a second-level analysis, to search for consistent significant effects across the whole group of participants. This random-effects analysis allows for inferences valid for the population from which the individuals have been recruited. However, any effects consistent by principle, that might be present for all, but that possibly

involve a large variance across listeners with respect to the exact position in the full brain volume, may become very weak or even stay hidden in this type of group analysis.

3. RESULTS

3.1 Diotic stimulus presentation

Figure 1 gives a summary of the results for the whole group of participants during diotic stimulation with sequences of noise bursts, fixed pitch notes, and melodies. The activation maps were superimposed on the mean of the normalized anatomical images from all fourteen participants. While the general activation in response to sound (blue) covers most of the top surface of the temporal lobes including Heschl's gyrus and part of the temporal plane in both hemispheres, the pitch specific activation (in red) is more restricted to the lateral edge of Heschl's gyrus bilaterally. The melody specific activation (depicted in green) appears somewhat more distributed over several structures adjacent to Heschl's gyrus, including areas on the temporal plane next to the presumed pitch center (symmetric, see left of Figure 1), and areas in right superior temporal sulcus (STS) and superior temporal gyrus (STG) on both sides.

Table 1 summarizes the mean coordinates across all 14 listeners for the main peaks in the activations maps for the contrast fixed pitch vs. noise, in case of diotic, i.e. binaural sound presentation. The bottom row in the table lists the corresponding coordinates as reported by Patterson et al. in 2002. The current results are in very good agreement with the previous study [9], with a deviation for the position of the mean peaks of only 3.5 mm in the left hemisphere and 6.6 mm in the right hemisphere.

Table 1. Mean activation peaks (plus standard deviation) across all individuals for the contrast between fixed pitch and noise conditions, for diotic stimulation. All coordinates are given in mm relative to the origin in the anterior commissure. The bottom row shows the results from the original study [9] for comparison.

	Left hemisphere			Right Hemisphere		
	<i>x</i>	<i>y</i>	<i>z</i>	<i>x</i>	<i>y</i>	<i>z</i>
This study	-53.7 (3.4)	-14.9 (6.0)	3.9 (3.6)	55.6 (4.4)	-9.1 (4.8)	4.5 (3.6)
Patterson et al., 2002	-55.3 (3.8)	-12.9 (4.6)	1.5 (2.9)	57.2 (2.5)	-8.8 (5.0)	-1.9 (3.3)

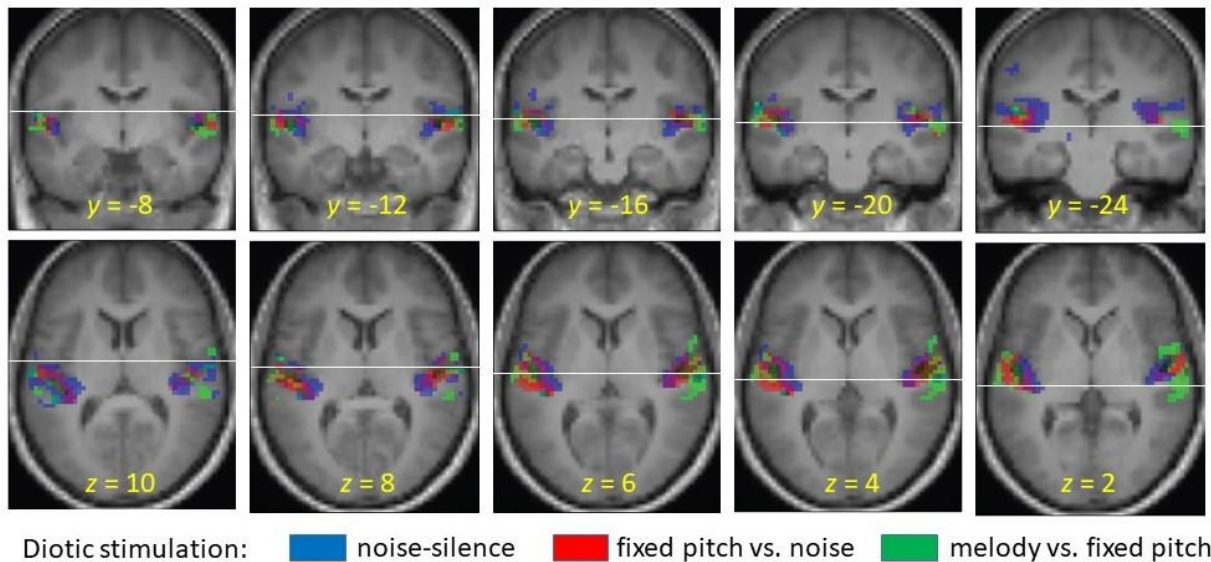


Figure 1. Activation maps for a group of fourteen listeners for the main contrasts noise vs. silence, fixed pitch vs. noise, and melody vs. fixed pitch, for diotic stimulus presentation. All maps have been superimposed on the average normalized structural image for this group. The activation maps have been gained from a 2nd-level random effects analysis across individual contrast images. The white line in each of the top panels indicates the position of the corresponding axial slices below, and vice versa. The respective coordinates are given in mm, relative to the origin at anterior commissure. Threshold for significance was set to $t = 3.85$ ($p < 0.001$), with the additional constraint of a minimum cluster size of eight connected voxels.

3.2 Monaural stimulus presentation

Figure 2 shows the effect of monaural stimulus presentation on the measured activation maps. The two monaural stimulus conditions (Fixed pitch and Melody) were combined for each ear of entry and contrasted with those from the respective other ear. These contrasts essentially illustrate the strength of the nerve fiber connections between the ears and the cortex, characterized by the crossed pathway on the way from cochlear nucleus (CN) to the upper brainstem (inferior colliculus, IC) and the cortex. Looking at the contrast Left vs. Right (left panel), there is no significant activation in the left auditory cortex. A similar observation holds for the corresponding opposite contrast (middle panel). Any contrast between the conditions with sound in general either to the left or to the right ear results in complete lateralization of the respective observable activation maps. There is not a single voxel in either hemisphere of the brain, that shows a bigger MR signal for ipsilateral than for contralateral stimulation.

The right panel illustrates the contrast between Right and Left ear stimulation at the level of the auditory brainstem. In this case, active clusters are found in the right CN (ipsilateral) but only in the left IC (contralateral), which is completely in line with the anatomy, with the first crossing on the pathway connecting CN and lateral lemniscus.

The middle and right panels in Fig. 3 now illustrate the contrasts between the conditions with melodies presented to one ear and sequences of fixed-pitch notes presented to the other ear, in comparison to the melody vs. fixed-pitch contrast for diotic stimulation (left panel). Activation around the surface of the temporal lobes, mainly in Heschl's gyrus (see coronal slices in the images at the top for the middle and right panel), is completely lateralized, as before (compare Fig. 2), reflecting the crossed projections from the periphery to primary auditory cortical areas. However, the melody specific areas outside Heschl's gyrus (the "green areas" from Fig. 1 and the left panel in Fig. 3) are always activated in both hemispheres in a similar manner, irrespective of the ear of entry.

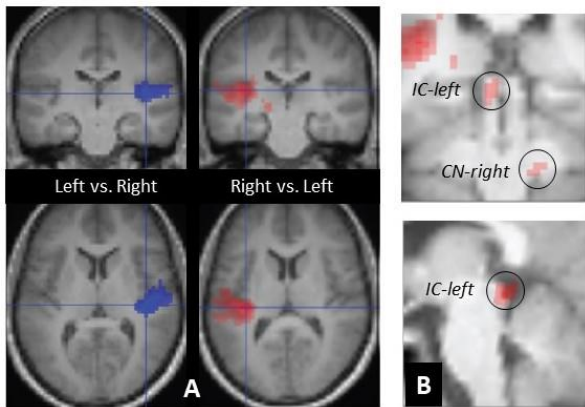


Figure 2. **A:** Activation maps for the contrasts between sound to the left only vs. sound to the right only (blue) and vice versa (Right only vs. Left only, red) at cortical level. Axial slices are at $z = 8$ mm to show the surface of the temporal lobes, including Heschl's gyrus with primary auditory cortex. Both monaural conditions, fixed pitch and melody, were combined. **B:** Activation map for Right vs. Left at the level of the auditory brainstem. Note that the activation includes the ipsilateral cochlear nucleus (CN), but the contralateral inferior colliculus (IC), indicating the crossed auditory pathway in between.

4. DISCUSSION

The main findings of the current study can be summarized as follows:

For diotic stimulus presentation, the results from this study are in very good agreement with the original report published in 2002 [9]. Any acoustic stimulus in contrast with a silent baseline condition will result in bilateral activation of both temporal lobes, which for simple acoustic stimuli like the repeated short bursts of random noise employed here as well as in [9] seems to be largely restricted to the surfaces of the temporal lobes, including Heschl's gyrus (HG) and the superior temporal gyrus (STG). Looking at the specific effect of repetition pitch realized in the iterated rippled noise stimuli employed here, when compared with the random noise condition, reveals pitch-specific activation mainly towards the lateral edge of HG, that is, beyond the primary auditory cortex located in the medial and central part of HG. Although the random effects analysis across the results from all 14 participants in this study suggests an activation pattern that is somewhat distributed over HG and the temporal plane, the quantitative

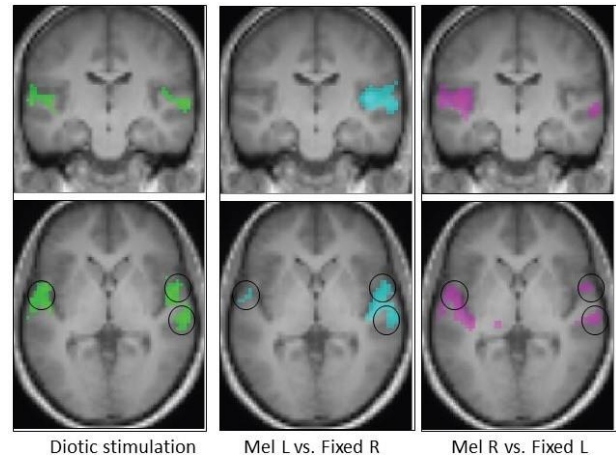


Figure 3. *Left panel:* contrast between melody and fixed pitch condition for diotic stimulus presentation (as in Fig. 1), for an axial slice (bottom) at the level of the anterior commissure ($z = 0$ mm), that is, below the surface of the temporal lobes. *Middle panel:* as before, but now also including the contrast left vs. right ear stimulation, that is: Melody left vs. Fixed pitch right. *Right panel:* as before, but now for the contrast Melody right vs. Fixed pitch left. Note that the melody specific activation in the axial slices appears always in the same positions in both hemispheres, irrespective of the particular ear of entry.

comparison of the main activation peak with the previously published results is surprisingly consistent, with a deviation of less than 4 mm in the left hemisphere and less than 7 mm in the right hemisphere for the group averages. This is another successful replication of the previous result, giving further evidence for the important role of lateral HG for the processing of acoustic stimuli with pitch. The results for diotic stimulation with respect to the activation specific to melodies, when contrasted with sequences of fixed-pitch notes, also essentially replicate what was reported before in 2002. With the group analysis, we find melody specific activation in both hemispheres, in regions beyond HG towards the poles of the temporal lobes, and, this mainly in the right hemisphere, also in regions below STG towards the sulcus and medial temporal gyrus. We interpret this as additional strong evidence for the hierarchical organization of temporal pitch processing.

The additional findings for monaural stimulus presentation suggest that the regions specifically responsive to changes of pitch reflect a higher processing stage. While a contrast

between left and right ear stimulation in general essentially reflects exactly what should be expected from the anatomy of the ascending auditory pathway, as illustrated in Fig. 2, the melody-specific activation seems to reflect a process of feature extraction from a highly processed, recoded sound representation, that integrates input from all pitch sensitive areas in the brain, irrespective of the particular ear of entry for the sound. In conclusion, it can be interpreted as representing a cognitive process beyond a purely sensory coding of stimulus properties, probably involving attention and individual musical aptitude.

5. ACKNOWLEDGMENTS

SU is supported by a travel grant from the DAAD.

6. REFERENCES

- [1] W.M. Hartmann, "Pitch, periodicity, and auditory organization," *J. Acoust. Soc. Am.*, vol. 100, pp. 3491-3502, 1996.
- [2] K. Krumbholz, R.D. Patterson, D. Pressnitzer, "The lower limit of pitch as determined by rate discrimination," *J. Acoust. Soc. Am.*, vol. 108, pp. 1170-1180, 2000.
- [3] D. Pressnitzer, R.D. Patterson, K. Krumbholz, "The lower limit of melodic pitch," *J. Acoust. Soc. Am.*, vol. 109, pp. 2074-2084, 2001.
- [4] W.A. Yost, "Pitch of iterated rippled noise," *J. Acoust. Soc. Am.*, vol. 100, pp. 511-518, 1996.
- [5] C.J. Plack, A.J. Oxenham, R.R. Fay, A.N. Popper (eds): *Pitch – neural coding and perception*. New York: Springer, 2005.
- [6] R.D. Patterson, "A pulse ribbon model of monaural phase perception," *J. Acoust. Soc. Am.*, vol. 82, pp. 1560-1586, 1987.
- [7] R.D. Patterson, J. Holdsworth, "A functional model of neural activity patterns and auditory images," in: *Advances in Speech, Hearing and Language Processing*, edited by W.A. Ainsworth, Volume 3B, JAI Press, pp. 547-563, 1996.
- [8] T.D. Griffiths, S. Uppenkamp, I.S. Johnsrude, O. Josephs, R.D. Patterson, "Encoding of the temporal regularity of sound in the human brainstem," *Nature Neuroscience*, vol. 4, pp. 633-637, 2001.
- [9] R.D. Patterson, S. Uppenkamp, I.S. Johnsrude, T.D. Griffiths, "The processing of temporal pitch and melody information in auditory cortex," *Neuron*, vol. 36, pp. 767-776, 2002.
- [10] S. Uppenkamp, A. Rupp, "Functional MR imaging of the processing of pitch changes in human listeners," in: *Fortschritte der Akustik - DAGA 2005*, Berlin: DEGA e.V., pp. 471-472, 2005.
- [11] D.A. Hall, I.S. Johnsrude, M.P. Haggard, A.R. Palmer, M.A. Akeroyd, A.Q. Summerfield, "Spectral and temporal processing in human auditory cortex," *Cereb. Cortex*, vol. 12, pp. 140-149, 2002.
- [12] K. Krumbholz, R.D. Patterson, A. Seither-Preisler, C. Lammertmann, B. Lütkenhöner, "Neuromagnetic evidence for a pitch processing center in Heschl's gyrus," *Cereb. Cortex*, vol. 13, pp. 765-772, 2003.
- [13] H. Penagos, J.R. Melcher, A.J. Oxenham, "A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging," *J. Neurosci.*, vol. 24, pp. 6810-6815, 2004.
- [14] A. Gutschalk, R.D. Patterson, M. Scherg, S. Uppenkamp, A. Rupp, "Temporal dynamics of pitch in human auditory cortex," *Neuroimage*, vol. 22, pp. 755-766, 2004.
- [15] I. Hertrich, K. Mathiak, H. Menning, W. Lutzenberger, H. Ackermann, "MEG responses to rippled noise and Huggins pitch reveal similar cortical representations," *Neuroreport*, vol. 16, pp. 193-196, 2005.
- [16] M. Chait, D. Poeppel, J.Z. Simon, "Neural response correlates of detection of monaurally and binaurally created pitches in humans," *Cereb. Cortex*, vol. 16, pp. 835-848, 2006.
- [17] D. Bendor, X. Wang, "The neuronal representation of pitch in primate auditory cortex," *Nature*, vol. 436, pp. 1161-1165, 2005.
- [18] M. Schönwiesner, R.J. Zatorre, "Depth electrode recordings show double dissociation between pitch processing in lateral Heschl's gyrus and sound onset processing in medial Heschl's gyrus," *Experimental Brain Research*, vol. 187, pp. 97-105, 2008.
- [19] J.O. Pickles: *An introduction to the physiology of hearing*, 4th ed., London: Emerald, 2012.
- [20] D.A. Hall, M.P. Haggard, M.A. Akeroyd, A.R. Palmer, A.Q. Summerfield, M.R. Elliott, E.M. Gurney, R.W. Bowtell, "'Sparse' temporal sampling in auditory fMRI," *Hum. Brain Mapp.*, vol. 7, pp. 213-223, 1999.