



Evaluating auditory stream segregation of SAM tone sequences by subjective and objective psychoacoustical tasks, and brain activity

Lena-Vanessa Dolležal¹, André Brechmann², Georg M. Klump^{1*} and Susann Deike²

¹ Animal Physiology and Behavior Group, Department for Neuroscience, School for Medicine and Health Sciences, Center of Excellence "Hearing4all," Carl von Ossietzky University Oldenburg, Oldenburg, Germany

² Special Lab Non-invasive Brain Imaging, Leibniz Institute for Neurobiology, Magdeburg, Germany

Edited by:

Elyse S. Sussman, Albert Einstein College of Medicine, USA

Reviewed by:

Andrew R. Dykstra, University of Heidelberg, Germany

Pierre Divenyi, Veterans Affairs Northern California Health Care System, USA

Makio Kashino, NTT Corporation, Japan

*Correspondence:

Georg M. Klump, Animal Physiology and Behavior Group, Department for Neuroscience, School for Medicine and Health Sciences, Center of Excellence "Hearing4all," Carl von Ossietzky University Oldenburg, Carl von Ossietzky Str. 9-11, D-26129 Oldenburg, Germany
e-mail: georg.klump@uni-oldenburg.de

Auditory stream segregation refers to a segregated percept of signal streams with different acoustic features. Different approaches have been pursued in studies of stream segregation. In psychoacoustics, stream segregation has mostly been investigated with a subjective task asking the subjects to report their percept. Few studies have applied an objective task in which stream segregation is evaluated indirectly by determining thresholds for a percept that depends on whether auditory streams are segregated or not. Furthermore, both perceptual measures and physiological measures of brain activity have been employed but only little is known about their relation. How the results from different tasks and measures are related is evaluated in the present study using examples relying on the ABA- stimulation paradigm that apply the same stimuli. We presented A and B signals that were sinusoidally amplitude modulated (SAM) tones providing purely temporal, spectral or both types of cues to evaluate perceptual stream segregation and its physiological correlate. Which types of cues are most prominent was determined by the choice of carrier and modulation frequencies (f_{mod}) of the signals. In the subjective task subjects reported their percept and in the objective task we measured their sensitivity for detecting time-shifts of B signals in an ABA- sequence. As a further measure of processes underlying stream segregation we employed functional magnetic resonance imaging (fMRI). SAM tone parameters were chosen to evoke an integrated (1-stream), a segregated (2-stream), or an ambiguous percept by adjusting the f_{mod} difference between A and B tones (Δf_{mod}). The results of both psychoacoustical tasks are significantly correlated. BOLD responses in fMRI depend on Δf_{mod} between A and B SAM tones. The effect of Δf_{mod} , however, differs between auditory cortex and frontal regions suggesting differences in representation related to the degree of perceptual ambiguity of the sequences.

Keywords: auditory scene analysis, amplitude modulation, temporal and spectral cues, time shift detection, BOLD response, fMRI

INTRODUCTION

In everyday life, the auditory system organizes acoustic signals based on similarities and differences in their sound features (Bregman, 1990). Especially in complex acoustic scenes, spectral or temporal stimulus parameters affect the grouping and segregation processes in assigning sounds to different sources. Sounds from one source are perceived as one coherent auditory stream. Studies of auditory stream segregation commonly applied the ABA- paradigm (e.g., Van Noorden, 1975; Moore and Gockel, 2002, 2012). The amount of stream segregation depends on the differences in sound features, i.e., the physical differences between A and B signals. It has been proposed that these differences will lead to the representation of the signals assigned to the separate streams by separate populations of neurons being differentially activated in time (e.g., Fishman et al., 2001; Elhilali et al., 2009). Furthermore, the separate

representation of the A and B signals can be observed already at the first stages of auditory processing (i.e., the cochlear nucleus) as Pressnitzer et al., 2008 demonstrated. Previous studies evaluated the auditory streaming percept and the neural mechanisms underlying the perceptual organization of sounds elicited by spectral (e.g., Van Noorden, 1975; Fishman et al., 2001, 2004; Bee and Klump, 2004, 2005; Deike et al., 2004, 2010; Micheyl et al., 2005; Micheyl and Oxenham, 2010) and temporal differences (Grimault et al., 2002; Vliegen and Oxenham, 1999; Roberts et al., 2002; Gutschalk et al., 2007; Itatani and Klump, 2011; Dolležal et al., 2012b) between A and B signals. These studies have provided evidence how auditory streaming is affected by a variety of simple features. Although applying a common paradigm, the outcome of these studies may also depend on the psychoacoustical task that was employed or on the measure that was used for evaluating the segregation of the streams. The majority of

the psychoacoustical tasks relied on a subjective perceptual judgment (e.g., Van Noorden, 1975; Vliegen and Oxenham, 1999; Grimault et al., 2002; Roberts et al., 2002; Micheyl et al., 2005; Gutschalk et al., 2007; Micheyl and Oxenham, 2010; Dolležal et al., 2012b). In the subjective task, subjects simply report their streaming percept. Few studies employed objective tasks (Van Noorden, 1975; Neff et al., 1982; Vliegen et al., 1999; Cusack and Roberts, 2000; Roberts et al., 2002; Micheyl and Oxenham, 2010; Thompson et al., 2011). In the objective task, the subject's perceptual threshold is determined using stimulus conditions in which threshold sensitivity is enhanced by one perceptual organization and hampered by the other (i.e., 1- and the 2-stream percept). Thus, in the objective task the streaming percept is inferred from the measured perceptual sensitivity. The different measures that were used range from the evaluation of perception to the assessment of the brain activity by applying invasive or non-invasive measurement techniques. Since few studies compared results obtained with different tasks (Micheyl and Oxenham, 2010) and measures (Gutschalk et al., 2007; Wilson et al., 2007), we have only little evidence how well these results are correlated.

Here, we investigate the correlation of the extent of stream segregation for sinusoidally amplitude modulated (SAM) A and B signals across two different psychoacoustical tasks and across two different measures (i.e., subjective psychoacoustical task and fMRI) providing a comprehensive approach to auditory stream segregation. The comparison across different psychoacoustical tasks involves an objective and a subjective task presenting signals with identical sound features to the same subjects. We propose that the thresholds obtained in the objective task are correlated with the subjective percept of stream segregation indicating that either task allows measuring the amount of perceptual stream segregation. Since any salient difference between sequential signals may elicit stream segregation (Moore and Gockel, 2002, 2012), we expect that a correlation will be found irrespective whether temporal or spectral cues can be utilized to differentiate between A and B signals. As is outlined in the methods below, SAM signals offer temporal, spectral, or both types of cues for stream segregation dependent on the modulation frequency (f_{mod}) and carrier frequency (f_c), respectively. By an appropriate choice of f_c and f_{mod} for the SAM tone stimulus sequences the amount of stream segregation between A and B SAM tones elicited by spectral cues and temporal cues can be varied (Dolležal et al., 2012a).

The comparison across measures includes the combination of the subjective task with fMRI. Previous human fMRI and MEG studies using either spectral (Deike et al., 2004, 2010; Gutschalk et al., 2005; Snyder et al., 2006; Wilson et al., 2007) or temporal (Gutschalk et al., 2007) differences between A and B stimuli consistently showed an increase of activity throughout the auditory cortex combined with a change of the dominant percept from 1-stream to 2-stream with increasing difference between stimuli. Based on these results we propose that A and B SAM tones show the same Δf_{mod} dependent activity in auditory cortex irrespective of the type of cue (i.e. spectral vs. temporal). The second goal pursued in obtaining fMRI activity measurements was to find further evidence of the specific involvement of regions

outside the auditory cortex in stream segregation which is still an open question in auditory streaming research. Cusack (2005), e.g., found that the intraparietal sulcus was differentially involved depending on the perceptual organization of physically identical stimuli in perceptual ambiguous sequences with a stronger BOLD activation for the segregated two-stream compared to the integrated one-stream percept. Such a segregation specific activation in the absence of physical differences was not found by Dykstra et al. (2011) when using intracranial EEG in neurosurgical patients with epilepsy. However, he observed a Δf dependent activity in middle temporal gyrus, pre- and post-central gyri, inferior and middle frontal gyri, and the supra-marginal gyrus. The human fMRI study by Kondo and Kashino (2009) found neural correlates of perceptual switching in the posterior insula, thalamus, and supra-marginal gyrus. Finally, some other studies did not describe evidence for the involvement of regions outside auditory areas in stream segregation (e.g., Wilson et al., 2007).

MATERIALS AND METHODS

SUBJECTS

Psychoacoustical measurements

Six human subjects (age 25–44 years, mean age 30 years, five females, including the first author) participated in two main experiments (Experiment 1 and 2; ABA- sequences) comparing subjective and objective psychoacoustical measures of stream segregation. All subjects had normal audiograms, with absolute pure tone thresholds <20 dB hearing level in the range from 0.25–10 kHz. Four of the subjects had previous experience with psychoacoustic experiments. In a control experiment (applying the conditions of Experiment 1 in B-only sequences) four of the subjects (age 25–42 years, mean age 30 years, three females, including the first author) participated. All experiments were undertaken with the understanding and written informed consent of each subject, following the Code of Ethics of the World Medical Association (Declaration of Helsinki). The experiments were approved by the local ethics committee of the University of Oldenburg. In addition to these psychoacoustical measurements in quiet, the subjective streaming percept was determined during fMRI (see fMRI measurements).

fMRI measurements

In Experiment 1, 13 subjects (age 20–31 years, mean age 26 years, five female) and in Experiment 2, 10 subjects (age 20–33 years, mean age 26 years, four female) participated. One subject participated in both experiments. Due to technical problems, psychophysical data of one subject in Experiment 2 are missing and only the data of 9 subjects were analyzed. All but one subject were right-handed (Edinburgh Handedness Inventory; laterality quotient $\geq +45$) and this one subject was ambidextrous (laterality quotient: 18). All showed a language laterality toward the left hemisphere tested as described in Bethmann et al. (2007). The subjects gave written informed consent to the study that was approved by the Ethics Committee of the University of Magdeburg. Five additional participants were excluded from the final analysis: one because of more than five missing responses and four because of head movements during

the fMRI-measurement that were stronger than 2.3 mm translation and/or 2.3° rotation.

APPARATUS, STIMULI, AND PROCEDURE

Stimuli

In the present study ABA- sequences (the dash indicates a silent interval of the same duration as the signal duration) were presented that consisted of fully sinusoidally amplitude modulated tones (SAM). ABA- sequences are commonly applied to determine the amount of stream segregation by varying the physical difference between A and B signals (Van Noorden, 1975). A and B signals with small physical differences are perceptually grouped into a single sequence (i.e., 1-stream percept) with a galloping rhythm (i.e., ABA-ABA-ABA-. . .), whereas A and B signals with large physical differences are perceptually segregated to two streams (i.e., 2-stream percept) with different isochronous rhythms (i.e., A-A-A-A-A-A-. . . and -B---B---B-. . .). For A and B signals with intermediate physical differences subjects may have an ambiguous percept, that is characterized by a switching between the 1- and the 2-stream percept (e.g., Moore and Gockel, 2002, 2012).

The SAM tones were digitally synthesized in Matlab (Version 7.1) at a sampling frequency of 44.1 kHz and produced by a Hammerfall DSP (Multiface II, RME). These signals (10 ms raised cosine rise/fall) had a duration of 125 ms and were presented at an overall presentation level of 70 dB SPL with a tone repetition time (TRT) of 250 ms. SAM tones have the advantage that the carrier frequency (f_c) and the modulation frequency (f_{mod}) can be adjusted in such a way that, depending on the parameter values and the auditory filter bandwidth (Kohlrausch et al., 2000), they provide either temporal, spectral, or both types of cues for stream segregation (Dolležal et al., 2012a). Dolležal et al. (2012a) used a computational model of the auditory periphery to calculate excitation pattern differences of A and B SAM tones and estimate spectral stream segregation thresholds based on these differences. If the observed thresholds were below the prediction based on spectral cues alone (i.e., could not be explained by spectral cues), they concluded that only temporal cues were relevant for the segregated percept. If the observed thresholds were similar or higher than the thresholds predicted on the basis of spectral cues, it was concluded spectral cues could provide a basis for perceptual stream segregation (for more details see Dolležal

et al., 2012a). **Table 1** summarizes the different parameter settings and highlights conditions in which spectral cues alone could explain stream segregation. For the remaining parameter settings, spectral cues are unlikely to explain stream segregation.

Here, ABA- sequences consisted of SAM tones that had the same carrier frequency (f_c) but different modulation frequencies (f_{mod}). Note the f_{mod} of the A SAM tones ($f_{mod A}$) was always lower than the f_{mod} of the B SAM tones ($f_{mod B}$). For the psychoacoustical tasks and the fMRI measurements, in the two different experiments the effect of $f_{mod A}$ (Experiment 1) or the effect of the f_c (Experiment 2) on stream segregation was analyzed. In Experiment 1 SAM tones had an f_c of 1 kHz and an $f_{mod A}$ of either 100 or of 300 Hz and in Experiment 2 SAM tones had an f_c of either 1 or of 4 kHz and an $f_{mod A}$ of 100 Hz. For each condition three f_{mod} differences between A and B SAM tones (Δf_{mod}) were chosen to evoke a 1-stream, a 2-stream and an ambiguous percept for the tested conditions, respectively (see **Table 1**). The value of Δf_{mod} was adjusted based on the study by Dolležal et al. (2012a). Dolležal et al. (2012a) also presented ABA SAM tone sequences, but they used either a TRT of 125 ms or a TRT of 375 ms for SAM tones of 125 ms duration. Based on their results the pre-set study chose for both experiments Δf_{mod} stimulus conditions that enable a comparison of stream segregation elicited by spectral and non-spectral cues. In Experiment 1 such a comparison can be made at the medium Δf_{mod} condition and in Experiment 2 at the large Δf_{mod} condition (**Table 1**; see Dolležal et al., 2012a). The results obtained in the present study were compared across two different psychoacoustical tasks and across two different measures (i.e., subjective psychoacoustical task and fMRI) for all Δf_{mod} stimulus conditions.

Psychoacoustical measurements

Psychoacoustical data were obtained in two different locations (Oldenburg and Magdeburg). In Oldenburg all subjects participated in the subjective and in the objective task in quiet in a sound-attenuating chamber (IAC, Industrial Acoustics Company, Mini 250). The stimuli were presented diotically with calibrated headphones (Sennheiser HDA 200). In Magdeburg, subjects participated in the subjective task during fMRI. Written instructions and additional verbal explanations, if necessary, were given to the subjects before the beginning of the tasks.

Table 1 | Stimulus conditions for Experiment 1 and Experiment 2.

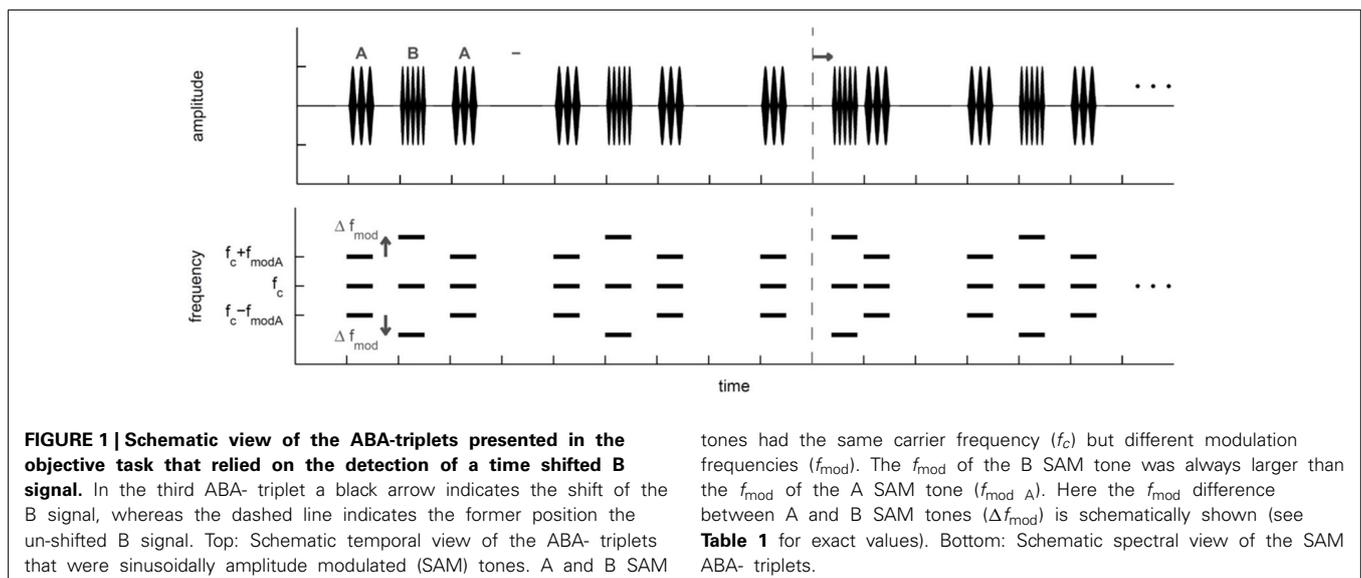
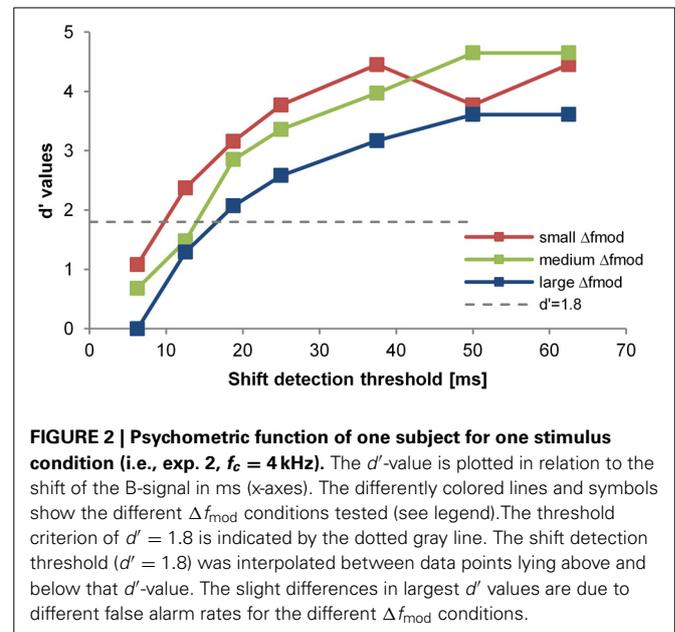
f_c [kHz]	$f_{mod A}$ [Hz]	Small Δf_{mod} [Hz]	Small Δf_{mod} [oct]	Medium Δf_{mod} [Hz]	Medium Δf_{mod} [oct]	Large Δf_{mod} [Hz]	Large Δf_{mod} [oct]
EXPERIMENT 1, EFFECTS OF $f_{mod A}$							
1	100	33	0.415	100	1.0	300	2.0
1	300	29	0.135	100	0.415	300	1.0
EXPERIMENT 2, EFFECTS OF f_c							
1	100	33	0.415	100	1.0	300	2.0
4	100	33	0.415	100	1.0	300	2.0

For each experiment, the carrier frequency (f_c), the modulation frequency of the A SAM tone ($f_{mod A}$) and the modulation frequency difference between A and B SAM tones (Δf_{mod}) are indicated. Δf_{mod} conditions that are printed in bold typeface show conditions in which stream segregation can be explained on the basis of spectral cues (Dolležal et al., 2012a) estimated from a model calculating excitation pattern differences between the A and the B SAM tones in the inner ear. Spectral cues provided by Δf_{mod} conditions that are printed in regular typeface are unlikely to be sufficient for stream segregation.

Objective task in quiet. Subjects started the experiment with the objective task in quiet. To measure objectively the perceptual segregation of the A and B SAM tones, subjects performed a shift detection task (Figure 1) in a Go/NoGo experiment determining the detection of a time shifted B SAM tone in the ABA-sequence. Thresholds obtained with the shift detection task should be smaller for ABA-sequences that are perceptually integrated into one stream than for ABA-sequences that are perceptually segregated to two streams (e.g., Van Noorden, 1975). In the present study, subjects listened to the presentation of a repeated ABA-triplet without a time shifted B SAM tone. Within 1 to 7 s (randomized time interval) after subjects started a trial by pressing a button on the touch screen either a forward shifted B SAM tone (Go-stimulus) replaced the regular B SAM tone or no replacement took place and a regular B SAM tone was presented (NoGo-stimulus, 30% of trials). If subjects detected the Go-stimulus in time (response latency < 1 s) by pushing a button on a touch screen, a correct response (hit) was registered and a green light flashed. If the subjects missed the Go-stimulus, a miss was recorded and the next trial was automatically initiated. The Go-response in this complex time-shift detection task could be based on the evaluation of the time interval between the A SAM tone and the successive B SAM tone or on the time interval between two sequential B SAM tones. Responses to NoGo-stimuli (false alarms) were registered too. Hit and false alarm rates were used to calculate the sensitivity measure d' (Green and Swets, 1966; see data analysis, psychoacoustics). For threshold estimation in each stimulus condition (Table 1) subjects had to complete a minimum of three sessions consisting of one obligatory training session and two subsequent test sessions (within each session a specific Go-stimulus was presented 10 times). A session lasted for about 20 min and consisted of eleven blocks of ten trials each. The first block of each session served as a warm-up block in which only the most salient Go-stimuli were presented. Each of the remaining ten blocks consisted of seven different Go-stimuli and three NoGo-stimuli that were presented in a random order. The Go-stimuli with a time shifted B SAM tone (step size 6.25

or 12.5 ms; i.e., 5 or 10% of the SAM tone duration) were chosen according to the method of constant stimuli. The range of the time shifts imposed on the B SAM tone was individually adjusted before each session to provide both sub-threshold and supra-threshold Go-stimuli. After each session a psychometric function was constructed relating the hits and misses of seven different Go-stimuli (different amounts of a time shifted B SAM tone) to d' -values (a measure of sensitivity for detecting the shift; see Figure 2). Between threshold sessions presenting different stimulus conditions a minimum pause of 5 min occurred. Within the objective task in quiet, the threshold estimation for the different stimulus conditions was randomized.

In addition to the objective task in quiet presenting ABA-sequences with time shifted B SAM tones, a control experiment



(for all stimulus conditions of Experiment 1) was conducted presenting B-only sequences. In B-only sequences only B SAM tones (omitting the A SAM tones) were presented (-B---B---B---) resulting in a TRT of 1000 ms. This experiment mimics a condition with a completely segregated percept, in which subjects solely rely on the stream of B SAM tones for the shift detection.

Subjective task in quiet. After performing in the objective task in quiet, subjects participated in the subjective task in quiet. Here, the same stimulus conditions as in the objective task were applied. ABA- sequences (15 s duration) of each stimulus condition were presented six times in randomized order. A pause of 45 s was introduced between the presentation of ABA- sequences of different f_c and $f_{\text{mod A}}$. After the presentation of each ABA- sequence subjects were instructed to indicate their percept (e.g., 1- or 2-stream percept) on a touch screen (Elo, 1542L, 15", Rear-Mount Touch-monitor). Then, the next ABA- sequence with another randomly chosen stimulus condition was initiated. Before starting the experiments in the subjective task in quiet, subjects attended a training session to familiarize with the task.

Subjective task during fMRI. During fMRI measurements, the subjects were presented with the same stimuli as in the subjective task in quiet. The duration of the stimulus sequences was increased to 16 s to adapt to the repetition time ($TR = 2000$ ms) of the functional echo planar imaging (EPI) sequence. Each of the three conditions (small, medium and large Δf_{mod}) were presented 10 times for each $f_{\text{mod A}}$ (100, 300 Hz) in Experiment 1 and for each f_c (1, 4 kHz) in Experiment 2, respectively, resulting in the presentation of 60 sequences per experiment. For each experiment, the order of the 60 sequences was pseudo-randomized with silence blocks of 16 s duration in between, which served as baseline condition. The stimuli were presented diotically via fMRI compatible headphones (Baumgart et al., 1998) at an individually adjusted, comfortable sound level, using Presentation (Neurobehavioral Systems Inc., San Francisco, USA). During the fMRI measurements, the subjects' heads were fixed with a cushion with attached earmuffs containing the headphones. Additionally, the subjects wore earplugs.

Prior to the fMRI measurements, the subjects received written instructions and additional verbal explanations if necessary. The subjects were asked to listen to the sound sequences and to indicate their percept at the end of each sequence by pressing the left button on a response panel with their right index finger when they perceived the SAM tones as one coherent stream, and the right button with their right middle finger when they perceived them as two separate streams. All button presses were recorded using Presentation (Neurobehavioral Systems Inc., San Francisco, USA) to test the perception of the SAM tone sequences under background scanner noise conditions. To familiarize the subjects with the sound sequences and the task, prior to the actual measurements, they were exposed to sequences, which most likely promote one or the other perceptual alternative, i.e., the 1-stream and the 2-stream percept, respectively.

fMRI measurements and data acquisition

The study was carried out on a 3 Tesla scanner (Siemens Trio; Erlangen, Germany) equipped with an eight channel head coil. A three-dimensional anatomical data set of the subject's brain (192 slices of 1 mm each) was obtained before the functional measurement. Additionally, before each functional run an Inversion-Recovery-Echo-Planar-Imaging (IR-EPI) with the identical geometry as in the functional measurement was acquired. Functional volumes were collected using a continuous EPI sequence (echo time $TE = 30$ ms; $TR = 2000$ ms; flip angle = 80° ; 32 slices; matrix size = 64×64 ; field of view (FOV) = 19.2 cm^2 , 3 mm isotropic resolution). The total experiment comprised 968 volumes scanned in 32 min 16 s.

DATA ANALYSIS

Psychoacoustical measurements

Psychoacoustical data were analyzed with repeated-measures analyses of variance (rmANOVAs, IBM SPSS Statistics Version 21.0). In all rmANOVAs, we report the F -values, the p -values and the partial η^2 , a non-additive value representing the "variance accounted-for" measure of the effect size, which can vary from 0 to 1 for the main effects. *Post-hoc* Tukey tests were Bonferroni corrected.

Objective task. For the threshold estimation of a stimulus condition data from two consecutive valid sessions in which thresholds differed by no more than 6.25 ms (i.e., 5% of the SAM tone duration) from each other were combined. A session was accepted as being valid based on two criteria: (1) Subject had a mean hit rate of 80% of the two easiest Go-stimuli (largest time shifts of the B SAM tone) and (2) their false alarm rate (NoGo-stimuli) was below 20%. Based on the rates of hits and misses, a psychometric function was constructed relating d' -values to each of the time shifts. By linearly interpolating between adjacent values of the psychometric function a shift detection threshold was determined as the time shift resulting in a d' -value of 1.8 (Green and Swets, 1966: **Figure 2**). To exclude training effects, the stimulus conditions of each experiment were randomized. Furthermore, after thresholds for all stimulus conditions were obtained subjects had to repeat the threshold measurement for the first condition of the series. If the new shift detection threshold differed by more than 6.25 ms from the shift detection threshold obtained in the first run subjects had to repeat measurements until the new threshold matched the threshold obtained in the first measurement (threshold difference ≤ 6 ms). In these cases the repeated shift detection threshold was taken for further analysis, discarding the previously measured threshold. In the rmANOVA, the shift detection thresholds were analyzed in relation to the stimulus condition (Δf_{mod}) and $f_{\text{mod A}}$ (Experiment 1) or f_c (Experiment 2).

Subjective task. For each subject and each condition the mean proportion of a 2-stream percept was calculated from the presentations of 6 (in quiet) or 10 sequences (during fMRI), respectively, per condition and then averaged across subjects. The proportion of a 2-stream percept in relation to the stimulus condition (Δf_{mod}) and $f_{\text{mod A}}$ (Experiment 1) or f_c (Experiment 2) was analyzed in a rmANOVA. The effect of the condition

of presentation (in quiet or during fMRI) on the proportion of a 2-stream percept was tested as between-subjects factor.

fMRI measurements

The functional data were analyzed using BrainVoyager™ QX (Brain Innovation, Maastricht, Netherlands). A standard sequence of pre-processing steps, such as 3D-motion correction, linear trend removal, and filtering with a high-pass of three cycles per scan was performed. The functional data sets were projected to the IR-EPI-images, co-registered with the 3D-data sets, and then transformed to Talairach space.

For each experiment separately, a conjunction analysis using a multi-subject random-effects general linear model (RFX-GLM) was performed to identify brain regions which showed positive deflections of the BOLD signal in at least one of the 3 conditions compared to the baseline ($t \geq 4.5$, $p < 0.002$ (uncorrected for multiple comparisons), cluster threshold: 108 mm^3) for each of the two stimulus variants:

Experiment 1: $f_{\text{mod A}} 100 \text{ Hz} > \text{baseline}$ AND $f_{\text{mod A}} 300 \text{ Hz} > \text{baseline}$,

Experiment 2: $f_c 1 \text{ kHz} > \text{baseline}$ AND $f_c 4 \text{ kHz} > \text{baseline}$.

The analysis included %-transformed functional data of all subjects and used the standard 2-gamma hemodynamic response function implemented in BrainVoyager™ QX. From the resulting clusters volumes-of-interest (VOIs) were defined. The BOLD responses of each VOI were subjected to repeated-measures analyses of variance (rmANOVAs) testing for the within factors condition (Experiment 1 and 2: small, medium and large Δf_{mod}), $f_{\text{mod A}}$ -variant (Experiment 1: 100, 300 Hz) and f_c -variant (Experiment 2: 1, 4 kHz). *Post-hoc* pair wise comparisons were performed using RFX-GLM analyses.

RESULTS

PSYCHOACOUSTICAL MEASUREMENTS IN QUIET AND DURING fMRI

The perceptual segregation of SAM tones was evaluated using either the subjective task (subjects directly reported their perceptual state in quiet or during fMRI) or the objective task that relied on the detection of a forward shifted B SAM tone within the ABA- sequence. In the first experiment the effect of $f_{\text{mod A}}$ was evaluated, whereas in the second experiment the effect of f_c was evaluated. Both, a variation of $f_{\text{mod A}}$ as well as a variation of f_c affects the representation of the SAM tones by temporal and/or spectral cues.

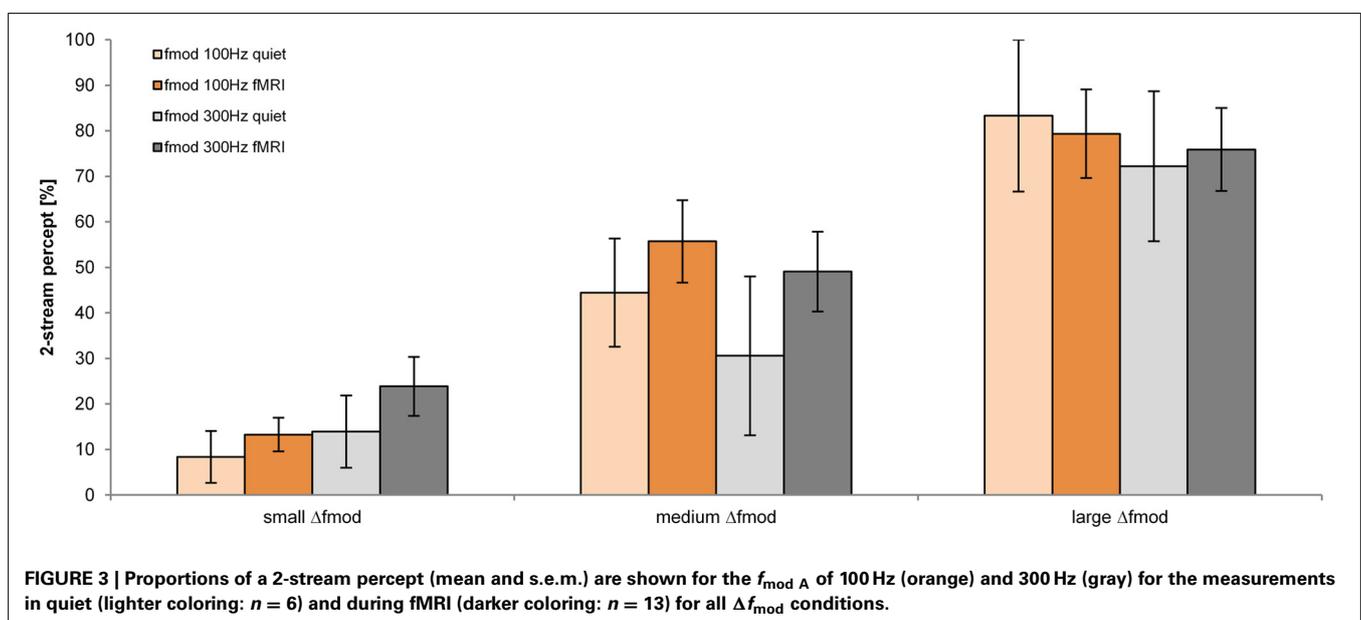
EXPERIMENT 1—THE EFFECT OF THE MODULATION FREQUENCY OF THE A SAM TONE ($f_{\text{mod A}}$)

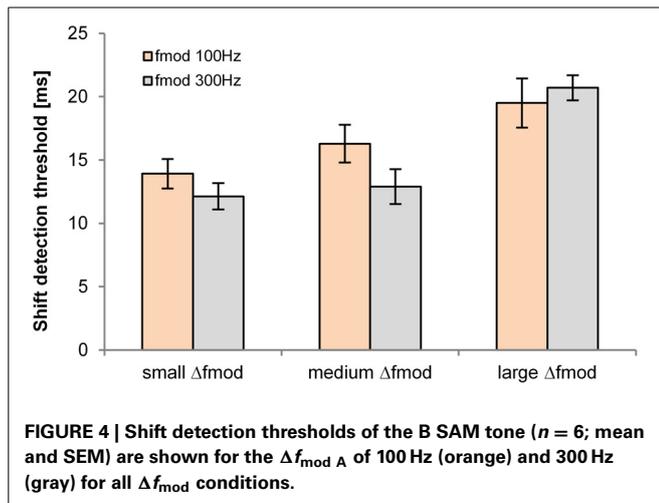
Subjective task

The proportion of a 2-stream percept depended significantly on the stimulus condition Δf_{mod} [$F_{(2, 34)} = 31.755$; $p < 0.001$, $\eta^2 = 0.651$]. The $f_{\text{mod A}}$ and the condition of presentation (in quiet and during fMRI) did not have a significant effect on the proportion of a 2-stream percept (Figure 3). Pair-wise comparisons showed a significant difference in the proportion of a 2-stream percept between all tested Δf_{mod} stimulus conditions (all $p \leq 0.003$). The mean proportion of a 2-stream percept increased significantly with increasing Δf_{mod} condition showing the least mean proportion of a 2-stream percept of 16.2% for ABA- sequences presented with the small Δf_{mod} condition. For ABA- sequences presented with the medium Δf_{mod} condition a mean proportion of a 2-stream percept of 47.7% was observed. The largest mean proportion of a 2-stream percept of 77.7% was observed for the large Δf_{mod} condition. No significant interaction was found.

Objective task in quiet

The shift detection threshold of the B signal was significantly affected by the Δf_{mod} stimulus condition [$F_{(2, 10)} = 38.795$; $p < 0.001$, $\eta^2 = 0.886$, Figure 4]. No significant main effect of $f_{\text{mod A}}$





on the shift detection threshold was observed. Pair-wise comparisons showed significantly higher shift detection threshold for the large (mean = 20.1 ms) than for the small (mean = 13 ms; $p = 0.001$) and medium Δf_{mod} condition (mean = 14.6 ms; $p = 0.001$). No significant difference between the shift detection threshold of the small and the medium Δf_{mod} condition was observed and no significant interaction was found.

Whether B SAM tones were presented by themselves (control experiment presentation of B-only sequences) or together with A SAM tones (only large Δf_{mod} condition of the main experiment, presentation of ABA- sequences) had a significant effect on the shift detection thresholds [$F_{(1, 3)} = 34.272$; $p = 0.01$, $\eta^2 = 0.920$]. Pair-wise comparisons showed significant higher mean shift detection thresholds for the control experiment (48.3 ± 2.4 ms) than observed in the large Δf_{mod} condition of the main experiment (mean = 20.1 ± 1.1 ms).

fMRI MEASUREMENTS

Table 2 lists all brain regions which were commonly activated or deactivated ($t = 4.5$, $p < 0.002$), respectively, by both $f_{\text{mod A}}$ in at least one of the three conditions (small, medium, and large Δf_{mod}) compared to the baseline condition.

In Experiment 1, the ANOVAs of BOLD responses within the respective VOIs revealed a main effect of Δf_{mod} condition in left Heschl's gyrus (HG) [$F_{(2, 24)} = 4.840$, $p = 0.017$] and left posterior cingulate gyrus (PCG) [$F_{(2, 24)} = 3.515$, $p = 0.045$]. In the left HG the BOLD response amplitude increased with increasing Δf_{mod} (see **Figure 5**). The *post-hoc* tests showed a significant difference between the small and the large Δf_{mod} condition ($t = 2.892$, $p = 0.013$) and a trend between the medium and the large Δf_{mod} condition ($t = 2.102$, $p = 0.057$). In left PCG, the *post-hoc* tests showed a significantly stronger negative deflection of the BOLD signal of the medium compared to the small Δf_{mod} condition ($t = 3.465$, $p = 0.005$).

In addition, in left and right HG [$F_{(1, 12)} = 22.800$, $p < 0.001$; $F_{(1, 12)} = 47.735$, $p < 0.001$], left and right insula [$F_{(1, 12)} = 6.685$, $p = 0.024$; $F_{(1, 12)} = 5.227$, $p = 0.041$], and the left posterior medial frontal cortex (pmFC) [$F_{(1, 12)} = 11.655$, $p = 0.005$] a main effect of $f_{\text{mod A}}$ was found with higher BOLD response

amplitudes during $f_{\text{mod A}}$ 300 Hz compared to $f_{\text{mod A}}$ 100 Hz stimulation (see **Figure 5**). There was no significant interaction of the factors Δf_{mod} condition and $f_{\text{mod A}}$.

EXPERIMENT 2—THE EFFECT OF THE CARRIER FREQUENCY (f_c)

Subjective task

The proportion of a 2-stream percept depended significantly on the Δf_{mod} condition [$F_{(2, 26)} = 51.595$; $p < 0.001$, $\eta^2 = 0.799$], on the f_c of the SAM tones [$F_{(1, 13)} = 11.623$; $p = 0.005$, $\eta^2 = 0.472$] and on the condition of presentation [in quiet or during fMRI; $F_{(1, 13)} = 8.168$; $p = 0.013$, $\eta = 0.386$; **Figure 6**]. Pair-wise comparisons showed a significant difference in the proportion of a 2-stream percept between all tested Δf_{mod} conditions (all $p \leq 0.009$). The proportion of a 2-stream percept increased significantly with increasing Δf_{mod} (mean percentage of a 2-stream percept for the small $\Delta f_{\text{mod}} = 11.9\%$, medium $\Delta f_{\text{mod}} = 44.0\%$ and large $\Delta f_{\text{mod}} = 86.9\%$). ABA- SAM tone sequences presented with the lower f_c of 1 kHz showed a significantly higher proportion of a 2-stream percept (50.5%) than SAM tones of the higher f_c of 4 kHz (44.7%). The proportion of a 2-stream percept measured in quiet was significantly smaller (mean = 35.2%) than the proportion of a 2-stream percept measured during fMRI (mean = 55.8%). The Two-Way interaction of the factors f_c and condition of presentation was significant ($p < 0.001$), showing a significant higher proportion of a 2-stream percept for the lower f_c of 1 kHz in quiet (mean = 45.4%) than for the higher f_c of 4 kHz in quiet (mean = 25.0%; $p = 0.006$), whereas the proportion of a 2-stream percept during fMRI was not affected by the f_c . No other interaction was significant.

Objective task in quiet

The detection threshold of the time shifted B SAM tone of the ABA- sequence was significantly dependent on the stimulus condition Δf_{mod} [$F_{(2, 10)} = 10.018$; $p = 0.004$, $\eta^2 = 0.667$, **Figure 7**]. No significant main effect of f_c on the shift detection threshold was observed. Pair-wise comparisons showed a significantly smaller shift detection threshold for the small (mean = 14.2 ms) than for the large Δf_{mod} stimulus condition (mean = 19.2 ms; $p = 0.01$). No significant difference between the shift detection threshold of the medium (mean = 15.9 ms) and the small and large Δf_{mod} was observed. No significant interaction was found.

fMRI MEASUREMENTS

Table 3 lists all brain regions which were commonly activated ($t = 4.5$, $p < 0.002$) by both f_c in at least one of the three conditions (small, medium, and large Δf_{mod}) compared to the baseline condition.

In Experiment 2, a main effect of condition was found in the left HG, the right superior temporal gyrus (STG), the left MedFG, and the right inferior parietal lobe (IPL) [$F_{(2, 18)} = 8.667$, $p = 0.002$; $F_{(2, 8)} = 19.634$, $p < 0.001$; $F_{(2, 18)} = 3.598$, $p = 0.048$; $F_{(2, 18)} = 3.501$, $p = 0.052$]. In left HG and right STG the same gradual increase in BOLD response amplitude with increasing Δf_{mod} was observed as in the left AC in Experiment 1 (see **Figures 5, 8**). *Post-hoc* tests in left HG and right STG revealed significant differences in BOLD responses between the small

Table 2 | Brain regions (BA-Brodmann area; x,y,z-Talairach coordinates) showing positive or negative deflections of the BOLD signal in at least one of the three Δf_{mod} conditions compared to the baseline ($t \geq 4.5$, $p < 0.002$) for each of the two $f_{\text{mod A}}$ (100, 300 Hz) tested in Experiment 1 and the results of ANOVAs within the resulting VOIs.

Brain region	VOI analysis (ANOVA)						
	Talairach coordinates				Main effect		
	BA	x	y	Z	Δf_{mod} condition	$f_{\text{mod A}}$	Volume (mm ³)
EXPERIMENT 1							
REGIONS OF ACTIVATION							
Left Hemisphere							
Heschl's gyrus	41	-48	-20	11	$F = 4.840$; $p = 0.017$	$F = 22.800$; $p < 0.001$	1307
Insula	13	-29	23	4	$F = 0.591$; $p = 0.562$	$F = 6.685$; $p = 0.024$	614
Putamen		-21	-1	8	$F = 0.101$; $p = 0.904$	$F = 0.972$; $p = 0.344$	150
Precentral gyrus	6	-38	2	30	$F = 0.091$; $p = 0.913$	$F = 2.591$; $p = 0.133$	207
Inferior parietal lobe	40	-36	-50	41	$F = 0.429$; $p = 0.656$	$F = 2.607$; $p = 0.132$	1047
Posterior medial frontal cortex	6	-5	8	46	$F = 0.048$; $p = 0.953$	$F = 11.655$; $p = 0.005$	389
Right Hemisphere							
Heschl's gyrus	41	53	-17	9	$F = 2.173$; $p = 0.136$	$F = 47.735$; $p < 0.001$	345
Superior temporal gyrus	22	62	-13	7	$F = 2.048$; $p = 0.151$	$F = 2.422$; $p = 0.146$	2357
Insula	13	32	21	2	$F = 1.804$; $p = 0.186$	$F = 5.227$; $p = 0.041$	390
Inferior parietal lobe	40	36	-48	40	$F = 0.375$; $p = 0.6$	$F = 1.017$; $p = 0.333$	1205
Middle frontal gyrus	9	47	15	26	$F = 0.215$; $p = 0.808$	$F = 0.863$; $p = 0.371$	738
Posterior medial frontal cortex	6	3	17	44	$F = 0.313$; $p = 0.734$	$F = 2.026$; $p = 0.180$	309
REGIONS OF DEACTIVATION							
Left Hemisphere							
Posterior cingulate gyrus	31	-48	-20	11	$F = 3.515$; $p = 0.045$	n.s.	138

Significant effects are highlighted in gray.

and the large Δf_{mod} condition ($t = 3.710$, $p = 0.005$; $t = 5.318$, $p < 0.001$) and between the small and the medium Δf_{mod} condition ($t = 5.727$, $p < 0.001$; $t = 5.929$, $p < 0.001$). In right STG, the large Δf_{mod} condition also resulted in a significantly stronger BOLD response than the medium Δf_{mod} condition ($t = 2.698$, $p = 0.024$). In left pMFC no gradual increase in BOLD response amplitude with increasing Δf_{mod} was observed. In contrast, the BOLD response of the medium Δf_{mod} condition was stronger than those of the small and the large Δf_{mod} condition (see **Figure 8**) with a significant difference between the medium and the small Δf_{mod} condition ($t = 2.258$, $p = 0.050$). The BOLD responses of the small and the large Δf_{mod} condition were very similar ($t = 0.643$, $p = 0.536$). *Post-hoc* testing in right IPL did not reach significance. No significant main effect of the f_c and no significant interaction of the factors condition and f_c were found.

CORRELATION BETWEEN TASKS AND MEASURES OF STREAM SEGREGATION

Correlation between tasks

For all subjects and the two main experiments the mean proportion of a 2-stream percept (subjective task in quiet) and the mean shift detection threshold (objective tasks in quiet) for the tested stimulus conditions were significantly correlated (Spearman's $\rho = 0.683$, $p = 0.042$, **Figure 9A**). The Spearman's non-parametric correlation coefficients for the single subject analyses were rather large ($\rho = 0.527$) for all but one subjects. Only for one subject the correlation reached a significant value ($p = 0.001$).

Correlation between measures

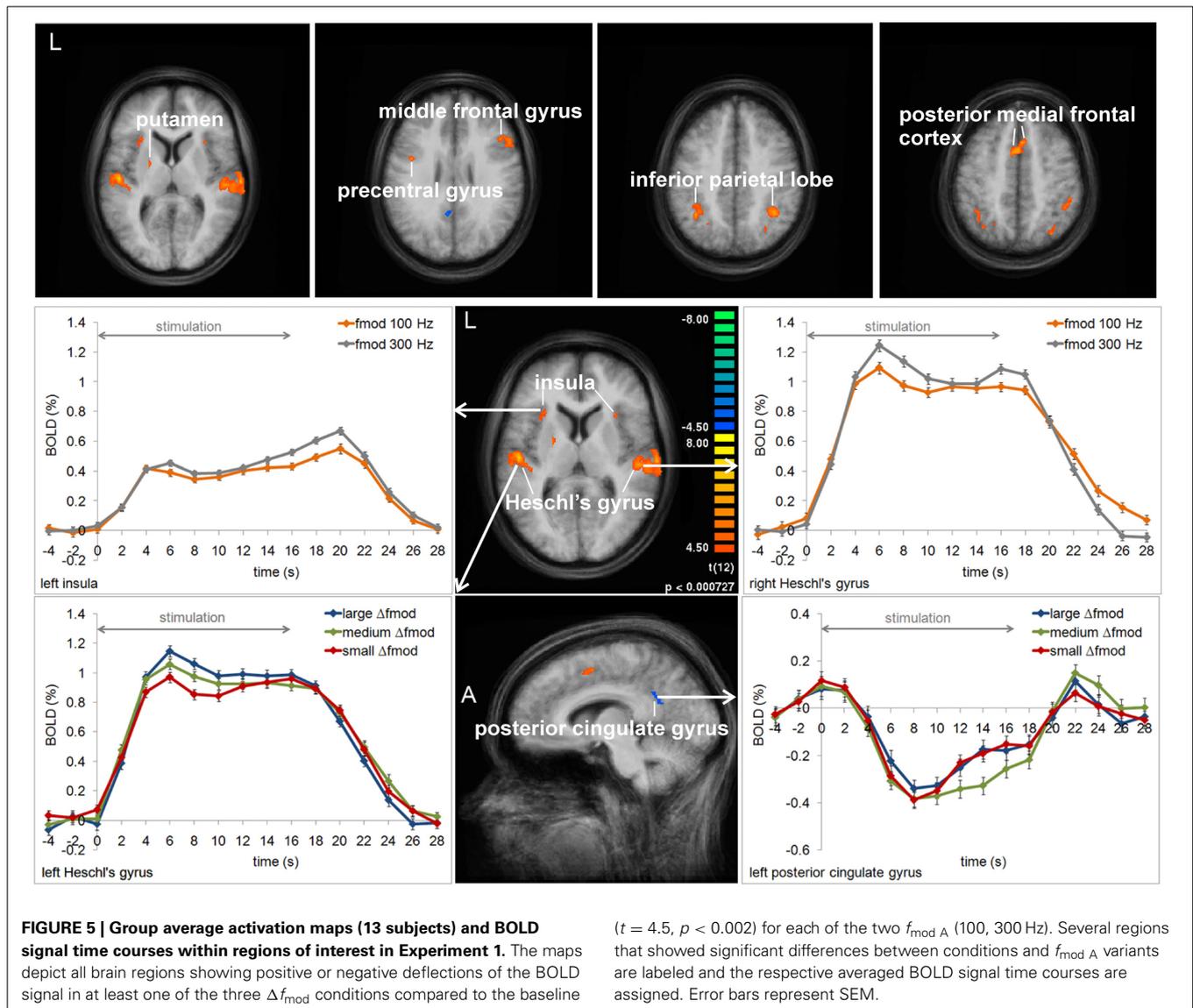
For the purpose of comparison and as an example, the proportion of a 2-stream percept for all tested Δf_{mod} stimulus conditions obtained during fMRI was related to the strength of BOLD responses (beta weights) in left Heschl's gyrus. Spearman's correlation of averaged group data did not reach significance ($\rho = 0.450$, $p = 0.224$, **Figure 9B**).

DISCUSSION

PSYCHOACOUSTICAL EVALUATION OF STREAM SEGREGATION BY SAM

The psychoacoustical results of both experiments and both the subjective and objective task show that an increasing Δf_{mod} between A and B SAM tones promotes stream segregation, being in agreement with the results of other psychoacoustical studies that evaluated stream segregation by either different SAM tones (e.g., Dolležal et al., 2012a; Szalárdy et al., 2013) or SAM noise bursts (Grimault et al., 2002).

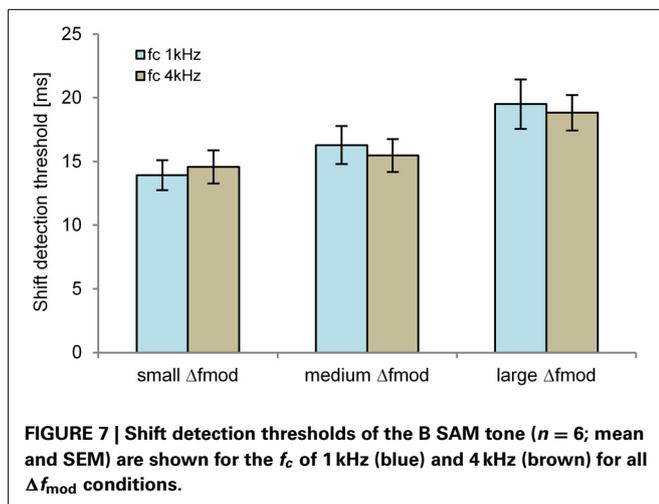
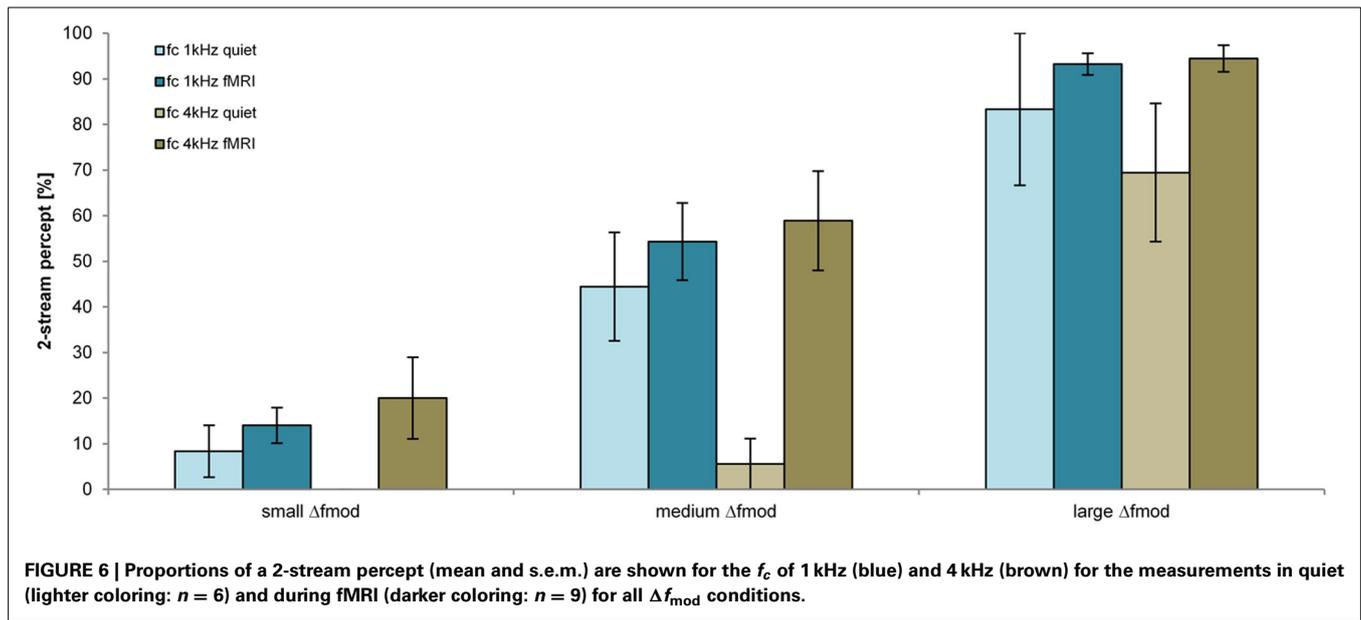
In the present study in Experiment 1 the subjective perception of stream segregation is not affected by the $f_{\text{mod A}}$ (100 and 300 Hz, respectively) of the SAM tones. Dolležal et al. (2012a), who presented sequences of SAM tones differing in multiple parameters in addition to $f_{\text{mod A}}$ and f_c [e.g., tone pattern (combinations of TRT and tone duration), modulation depth and presentation time] and used more steps of Δf_{mod} , however, observed an increasing proportion of a 2-stream percept for increasing $f_{\text{mod A}}$ (30, 100, and 300 Hz). This difference between the two



studies could be attributed to the differences in the range of $f_{\text{mod A}}$ and Δf_{mod} that was larger in the previous study by Dolležal et al. (2012a). When comparing the proportion of a 2-stream percept of the medium Δf_{mod} condition, that was explicitly chosen to compare stream segregation of temporal ($f_{\text{mod A}} = 100$ Hz) vs. spectral ($f_{\text{mod A}} = 300$ Hz) cues (Table 1), spectral cues appear not to further stream segregation more than temporal cues. Next to the evaluation of the subjective streaming percept the present study also applied an objective task of stream segregation to the same stimulus conditions to be able to directly compare the streaming percept across both psychoacoustical tasks. The shift detection thresholds obtained with the objective task increased with increasing Δf_{mod} between A and B SAM tones. Such an increase in the shift detection threshold with increasing feature differences between A and B signals has been observed in other studies that also presented time shifted signals in an objective task using a range of different features (frequency differences: Van Noorden, 1975; Neff et al., 1982; Cusack and Roberts, 2000;

Micheyl and Oxenham, 2010; Thompson et al., 2011; differences in the starting phases of frequency components: Roberts et al., 2002; differences in fundamental frequencies: Vliegen et al., 1999). Furthermore, Divenyi and Danner (1977) also observed a sizable deterioration of the discrimination performance if the signals were made very dissimilar from each other (e.g., in frequency or intensity) even though they did not employ a paradigm that led to a streaming percept.

We also applied the shift detection task in an ABA- sequence with omitted A signals (-B---B---B---) to determine the shift detection threshold in a condition providing no temporal reference to A signals. We observed higher shift detection thresholds in the B-only condition than for the large Δf_{mod} condition of ABA- sequences with A and B signals. That difference in threshold may indicate that even in sequences with well segregated A and B signals the A signal can provide support to the detection of the time shift of the B signal. If subjects would have solely relied on the B SAM tones for their performance in both the B only



condition and in the ABA- condition the thresholds should be alike.

In Experiment 2 the subjective perception of stream segregation was affected by the Δf_{mod} condition and by f_c (1 and 4 kHz, respectively) of the SAM tones when analyzing the subjective data from fMRI and those obtained in quiet together. The effects of the f_c and Δf_{mod} and their interaction was also observed by Dolležal et al. (2012a) who reasoned that the difference in the proportion of a 2-stream percept may be due to the excitation pattern differences between A and B signals being assessed by the auditory system. In the present analysis, a higher proportion of a 2-stream percept was observed for the lower f_c of 1 kHz than for the higher f_c of 4 kHz. At a f_c of 1 kHz at least in the large Δf_{mod} condition spectral cues provided for stream segregation in addition to the temporal cues that were also the prominent cue for ABA- SAM tones presented at a f_c of 4 kHz. Thus, the spectral excitation pattern difference available for the lower f_c of 1 kHz

providing additional cues to stream segregation may be the cause for the higher amount of a 2-stream percept in that condition. The significant interaction between the condition of presentation (quiet, during fMRI) and f_c , however, indicates that responses differed between both presentation conditions. The effect of f_c was only prominent in quiet conditions and not in the noisy fMRI condition that may have precluded the use of excitation pattern differences. The subjective segregation percept in the noisy fMRI condition match the pattern of BOLD responses (see below). If we focus on the large Δf_{mod} condition that allows comparing the amount of stream segregation elicited by spectral vs. temporal cues, we find no significant difference indicating that both type of cues have the potential to elicit the percept of well segregated streams.

In general, the proportion of a 2-stream percept was smaller for subjects that have been tested in quiet, than for subjects that have been tested in scanner noise during fMRI measurements. Especially in the medium Δf_{mod} condition we observed a small amount of stream segregation that was less than expected on the basis of the previous measurements (Dolležal et al., 2012a). A similar difference in the proportion of a 2-stream percept has been observed in Experiment 1, but it did not reach significance. Wilson et al. (2007) also compared the streaming perception of subjects in quiet and during fMRI. Their results show a non-significant but higher proportion of a 2-stream percept for subjects tested during fMRI than in the quiet booth revealing a tendency that is comparable to the results of the present study. Dolležal et al. (2012a) also observed a higher proportion of a 2-stream percept in pink noise than in quiet. A general explanation for the observed effect, however, cannot be provided.

When presenting the stimulus conditions of Experiment 2 in the objective task no effect of f_c on the shift detection threshold can be observed whereas an effect of Δf_{mod} remained. In the subjective task the effect size of Δf_{mod} was considerably larger than the effect size for f_c . Since the objective task will lead to better

Table 3 | Brain regions (BA-Brodmann area; x,y,z-Talairach coordinates) showing positive deflections of the BOLD signal in at least one of the three Δf_{mod} conditions compared to the baseline ($t = 4.5$, $p < 0.002$) for each of the two f_c (1, 4 kHz) tested in Experiment 2 and the results of ANOVAs within the resulting VOIs.

Brain region	VOI analysis (ANOVA)						
	Talairach coordinates				Main effect		
	BA	x	y	Z	Δf_{mod} condition	$f_{\text{mod A}}$	Volume (mm ³)
EXPERIMENT 2							
REGIONS OF ACTIVATION							
Left Hemisphere							
Heschl's gyrus	41	-44	-24	12	<i>F</i> = 8.667; <i>p</i> = 0.002	<i>F</i> = 0.016; <i>p</i> = 0.900	540
Inferior parietal lobe	40	-33	-52	40	<i>F</i> = 1.014; <i>p</i> = 0.383	<i>F</i> = 0.016; <i>p</i> = 0.902	126
Posterior medial frontal cortex	6	-3	14	46	<i>F</i> = 3.598; <i>p</i> = 0.048	<i>F</i> = 1.000; <i>p</i> = 0.343	257
Right Hemisphere							
Superior temporal gyrus	22	51	-13	10	<i>F</i> = 19.634; <i>p</i> < 0.001	<i>F</i> = 3.177; <i>p</i> = 0.108	592
Insula	13	36	19	4	<i>F</i> = 0.859; <i>p</i> = 0.440	<i>F</i> = 0.483; <i>p</i> = 0.504	113
Inferior parietal lobe	40	38	-45	46	<i>F</i> = 3.501; <i>p</i> = 0.052	<i>F</i> = 1.650; <i>p</i> = 0.231	541

Significant effects are highlighted in gray.

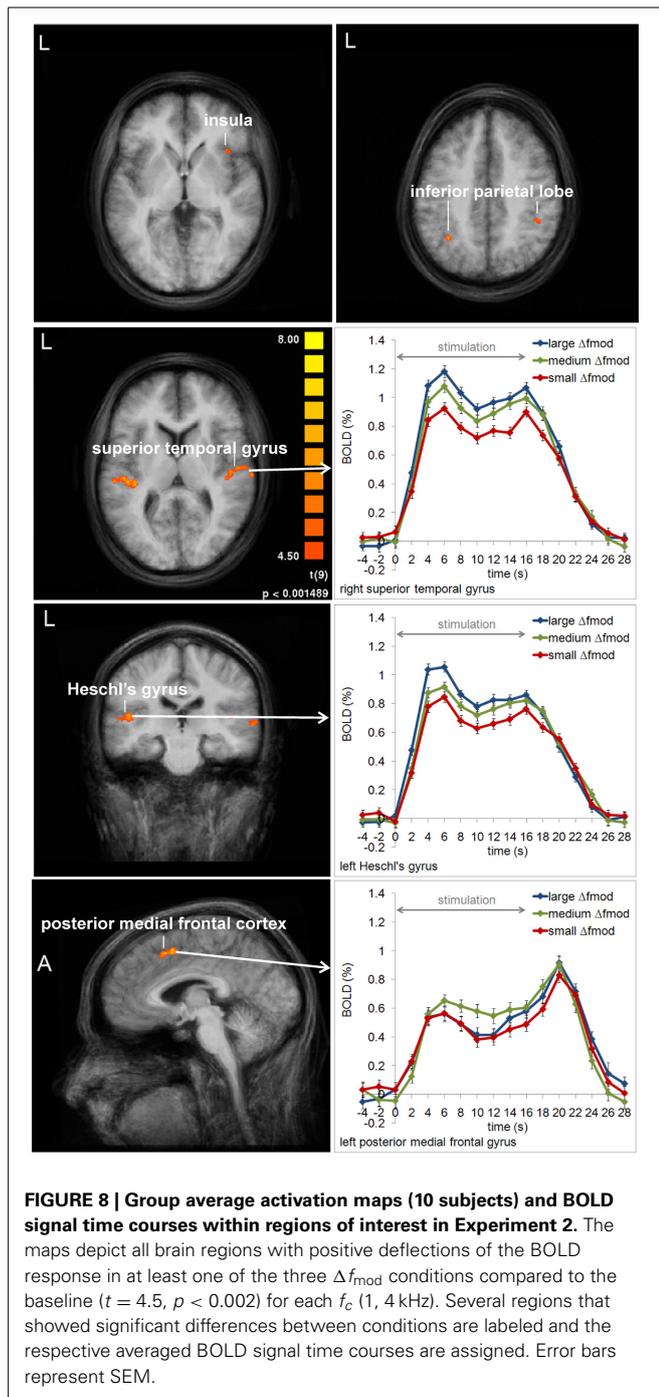
thresholds if the subjects integrate A and B signals into a single stream (e.g., Van Noorden, 1975; Neff et al., 1982; Vliegen et al., 1999; Cusack and Roberts, 2000; Roberts et al., 2002; Micheyl and Oxenham, 2010; Thompson et al., 2011), they may be inclined to integrate more than in a subjective evaluation of the stimuli. This may reduce smaller effects of the subjective task to non-significance in the objective task.

BOLD ACTIVITY DURING STREAM SEGREGATION BY SAM TONES

Corresponding to the psychoacoustical results, BOLD activity in auditory cortex regions depended on the Δf_{mod} between A and B SAM tones. With increasing Δf_{mod} the dominant percept changed from a 1-stream to a 2-stream and the BOLD response amplitudes gradually increased. The results of Experiment 1 and 2 differ, however, in that the Δf dependent effect was observed only in left auditory cortex in Experiment 1 and in both auditory cortices in Experiment 2. Previous human imaging studies on stream segregation found either an involvement of both auditory cortices (e.g., Gutschalk et al., 2007; Wilson et al., 2007) or a specific involvement of the left auditory cortex (Deike et al., 2004, 2010). Deike et al. suggested that the involvement of the left hemisphere was caused by the specific demands on sequential analysis in the active stream segregation task. Even though the present experiments require the sequential analysis of the sound sequences, the subjects were not forced to actively group the sounds into one or the other perceptual organization but had to monitor their spontaneous perception. Therefore, one might rather suggest a stimulus driven representation of Δf_{mod} in both auditory cortices and the failure to observe this in right auditory cortex in Experiment 1 might simply be explained by statistical thresholding.

The Δf_{mod} dependent effect in auditory cortex regions was observed for all stimulus parameters and thus, irrespective as to whether SAM tones provide spectral, temporal or both types of cues. Several human imaging studies have described increasing neural activity throughout the auditory cortex for both differences in spectral (Deike et al., 2004, 2010; Gutschalk et al.,

2005; Snyder et al., 2006; Wilson et al., 2007) and in temporal (Gutschalk et al., 2007) properties between A and B signals in streaming sequences. Hence, our finding of increasing BOLD response amplitudes in auditory cortex regions with increasing Δf_{mod} between SAM tones is consistent with previous studies. Electrophysiological recording studies in animals using pure-tone paradigms suggested that frequency selectivity of tonotopically organized neurons in primary auditory cortical fields in combination with forward suppression leads to separate representations of A and B tones that contribute to the percept of two separate streams (Fishman et al., 2001, 2004; Kanwal et al., 2003; Bee and Klump, 2004, 2005; Micheyl et al., 2005). With increasing frequency separation between tones the populations of active neurons become more disjointed, leading to decreasing suppression between successive tones. It was supposed that this decrease in suppression causes the larger summed activity in auditory cortex measured using fMRI, EEG, or MEG (Gutschalk et al., 2005; Snyder et al., 2006; Wilson et al., 2007). Using harmonic tone complexes with only unresolved harmonics Gutschalk et al. (2007) suggested that suppression also accounts for the interaction of sounds with differences in temporal properties. For SAM tones Bartlett and Wang (2005) found that neurons in marmoset monkey auditory cortex show significant forward suppression of the preceding to the following SAM tone. Similarly, Itatani and Klump (2009), who used the same ABA- paradigm as in the present study and tested a large parameter space of SAM tones, observed forward suppression in multiunit responses of the auditory forebrain of awake European starlings. Related to this potential common cortical mechanism underlying stream segregation on temporal and spectral properties of sounds one may further ask the question of pitch representation at the cortex. In the present study, two stimulus conditions (Experiment 1: medium Δf_{mod} , Experiment 2: large Δf_{mod}) provided a direct comparison between spectral and temporal pitch cues on which stream segregation was based and we did not find any cortical region which showed a significant difference in BOLD responses

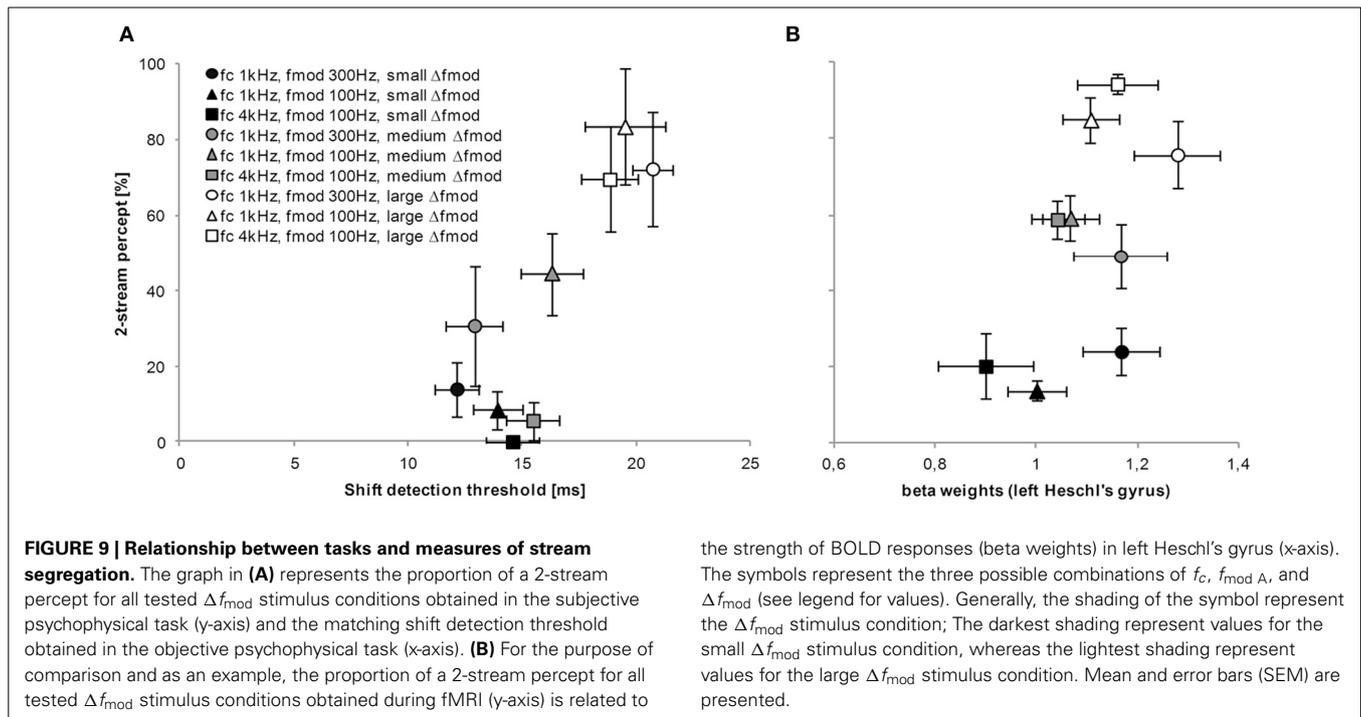


between both cues in this comparison. This finding is consistent with the results by Hall and Plack (2009) who tested a range of pitch-evoking stimuli with different spectral, temporal, and binaural characteristics and did not find any differentiated activation within auditory cortex regions. Although differing in anatomical location, there is supporting evidence for a cue independent common pitch region in auditory cortex coming from the neurophysiological study by Bendor and Wang (2005) who found pitch-selective neurons near the anterolateral low-frequency border of

the primary auditory cortex field A1 in marmoset monkeys. At the same time, the underlying mechanism for pitch coding in this region was found to depend both on the temporal and spectral characteristics of the sounds (Bendor et al., 2012).

In Experiment 1, BOLD responses in left and right Heschl's gyrus depended on the $f_{\text{mod A}}$ of SAM tones with higher BOLD response amplitudes for the higher $f_{\text{mod A}}$ of 300 Hz compared to the smaller one of 100 Hz. As SAM tones are characterized by three spectral peaks, i.e., the central peak representing the f_c and the two sidebands (upper: $f_c + f_{\text{mod A}}$, lower: $f_c - f_{\text{mod A}}$), the stronger responses for the higher $f_{\text{mod A}}$ of 300 Hz might be explained by broader spectral excitation. In addition, higher BOLD response amplitudes for the higher $f_{\text{mod A}}$ were also observed in left and right insula and in the left medial part of Brodmann area 6 comprising the supplementary motor area (SMA). The involvement of these areas might be thought in the context of specific task demands other than motor processing in which both have a primary function. Specifically, the insula cortex has a role in different auditory processes, such as allocating auditory attention, temporal processing, phonological processing, and visual-auditory integration (for review, see Bamjiou et al., 2003). The SMA is described as a part of the larger functional unit of posterior medial frontal cortex (pmFC) which has a function in cognitive control and particularly in performance monitoring including monitoring of response conflicts and decision uncertainty (for review, see Ridderinkhof et al., 2004). As the subjects' task was to assign their perception to one of the two perceptual alternatives, the stronger activity for the $f_{\text{mod A}}$ of 300 Hz in the pmFC might reflect the monitoring of a response conflict or uncertainty in perceptual decision. In the same way, Tregellas et al. (2006) observed in the pmFC and the insular/opercular cortex an increase in BOLD activity in a "difficult" compared to an "easy" auditory temporal processing task. In their study, the subjects had to discriminate the duration of the second tone within pairs of tones and the task difficulty was adapted by varying duration differences between tones. Increasing BOLD activity in the anterior insula bilaterally with increasing task demands were also observed in a pitch discrimination and a n-back pitch memory task (Rinne et al., 2009). In the present study the task required sequence processing and rhythmic pattern perception, namely comparing the galloping ABA- rhythm (1-stream percept) to the two different isochronous rhythms (A-A-A-... and -B--B-...; 2-stream percept). Although the proportion of 2-stream perception is very similar across conditions between both $f_{\text{mod A}}$ variants one might suppose that the perceptual decision might be more difficult for the higher $f_{\text{mod A}}$ of 300 Hz because of specific sound qualities (e.g., timbre) other than pitch. Corresponding to this, one might suggest that the stronger BOLD response in auditory cortex for the higher $f_{\text{mod A}}$ of 300 Hz also reflects the task difficulty. This notion finds support in human imaging studies providing evidences that even in sensory areas the activation can be modulated by task difficulty (Gerlach et al., 1999; Brechmann and Scheich, 2005; Reiterer et al., 2005; Harinen and Rinne, 2013).

In both experiments, cortical regions outside the auditory cortex were found which showed specific activity for medium Δf_{mod} between SAM tones compared to the small and the large Δf_{mod} 's. In particular, the left pmFC (Experiment 2) and the left posterior



cingulate gyrus (PCG) (Experiment 1) showed stronger positive and negative deflections of the BOLD signal, respectively, for the medium Δf_{mod} and very similar smaller BOLD responses for the two other conditions. This activation pattern is very different from the gradual increase in activation with increasing Δf_{mod} that was observed in auditory cortex regions. Whereas the activation gradient in auditory cortex rather reflects the physical differences between conditions, the BOLD responses in the pMFC and the PCG might rather be related to perceptual decision. As already mentioned above, the pMFC has a cognitive function in response conflicts and decision uncertainty. This is particularly the case in the ambiguous perceptual region where both perceptual alternatives are possible and compete with each other. Thus, the stronger BOLD response for ambiguous sequences in the pMFC might be explained by response conflicts and/or decision uncertainty. Similarly, response conflicts or decision uncertainty are equivalent to imposing higher task demands that might explain the stronger deactivation for ambiguous sequences in the PCG which is a part of the “task negative” default mode network showing decreasing activity with increasing task demands (Raichle et al., 2001; Corbetta and Shulman, 2002; Fox et al., 2005; Dosenbach et al., 2007).

Our fMRI results can be summarized as follows. In auditory cortex stream segregation on SAM tones showed the same Δf dependent BOLD responses as other streaming stimuli. In contrast, BOLD activity in regions outside the auditory cortex rather appear to reflect the perceptual decision and specifically the higher task demands caused by specific stimulus characteristics or by perceptual ambiguity leading to response conflicts and decision uncertainty, respectively. The involved regions differ from those observed in other studies and we did not find significant activation in any of the regions reported in Cusack (2005)

(intraparietal sulcus), Kondo and Kashino (2009) (Thalamus), and Dykstra et al. (2011) (e.g., middle temporal and frontal gyri). This might be explained by general differences in the approaches: Cusack (2005) and Kondo and Kashino (2009) examined ambiguous streaming sequences to find correlates of different perceptual organizations and perceptual switches, respectively, whereas the present study examined stream segregation across the domains of perceptual dominance and ambiguity by varying the stimulus parameters. The study by Dykstra et al. (2011) also compared different Δf conditions and found that the middle temporal and frontal gyri showed the same increase in neural activity with increasing Δf as the auditory cortex. They, however, did not observe a specific response for ambiguous stimuli. This discrepancy must be resolved in future studies.

COMPARISON ACROSS TASKS AND MEASURES

A direct comparison across psychoacoustical tasks of stream segregation showed a correlation across all subjects and experiments (Figure 9A). The results of the objective task mirror the results obtained by the subjective task, thus the shift detection threshold as well as the proportion of a 2-stream percept increased with increasing Δf_{mod} between A and B SAM tones. Such a correlation reveals that both the proportion of a 2-stream percept (subjective task) as well as the shift detection threshold (objective task) can represent the amount of stream segregation. These results are in agreement with a study by Micheyl and Oxenham (2010) who presented pure tones in ABA- sequences with frequency differences between A and B tones and also correlated the proportion of a 2-stream percept with the shift detection experiment. A comparison across measures (subjective streaming percept and BOLD responses (beta weights) in left Heschl's gyrus) did not show a significant correlation even though a

relatively high Spearman's rho was observed for the mean values of both measures (**Figure 9B**). In the exemplary figure of the correlation of humans perception and BOLD responses in left Heschl's gyrus a trend similar to the one observed in the figure of the correlation across psychoacoustical tasks (see **Figure 9A**) can be observed, showing an increasing proportion of a 2-stream percept with increasing beta weights measured in BOLD responses.

ACKNOWLEDGMENTS

This study was supported by the DFG (SFB TRR 31, GRK 591). We thank Rainer Beutelmann, Holger Dierker, Monika Dobrowolny, and Antje Schasse for assistance.

REFERENCES

- Bamiou, D.-E., Musiek, F. E., and Luxon, L. M. (2003). The insula (Island of Reil) and its role in auditory processing: Literature review. *Brain Res. Rev.* 42, 143–154. doi: 10.1016/S0165-0173(03)00172-3
- Bartlett, E. L., and Wang, X. (2005). Long-lasting modulation by stimulus context in primate auditory cortex. *J. Neurophysiol.* 94, 83–104. doi: 10.1152/jn.01124.2004
- Baumgart, F., Kaulisch, T., Tempelmann, C., Gaschler-Markefski, B., Tegeler, C., Schindler, F., et al. (1998). Electrodynamical headphones and woofers for application in magnetic resonance imaging scanners. *Med. Phys.* 25, 2068–2070. doi: 10.1118/1.598368
- Bee, M. A., and Klump, G. M. (2004). Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *J. Neurophysiol.* 92, 1088–1104. doi: 10.1152/jn.00884.2003
- Bee, M. A., and Klump, G. M. (2005). Auditory stream segregation in the songbird forebrain: effects of time intervals on responses to interleaved tone sequences. *Brain Behav. Evol.* 66, 197–214. doi: 10.1159/000087854
- Bendor, D., Osmanski, M. S., and Wang, X. (2012). Dual-pitch processing mechanisms in primate auditory cortex. *J. Neurosci.* 32, 16149–16161. doi: 10.1523/JNEUROSCI.2563-12.2012
- Bendor, D., and Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature* 436, 1161–1165. doi: 10.1038/nature03867
- Bethmann, A., Tempelmann, C., De Bleser, R., Scheich, H., and Brechmann, A. (2007). Determining language laterality by fMRI and dichotic listening. *Brain Res.* 1133, 145–157. doi: 10.1016/j.brainres.2006.11.057
- Brechmann, A., and Scheich, H. (2005). Hemispheric shifts of sound representation in auditory cortex with conceptual listening. *Cereb. Cortex* 15, 578–587. doi: 10.1093/cercor/bhh159
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.
- Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755
- Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *J. Cogn. Neurosci.* 17, 641–651. doi: 10.1162/0898929053467541
- Cusack, R., and Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Percept. Psychophys.* 62, 1112–1120. doi: 10.3758/BF03212092
- Deike, S., Gaschler-Markefski, B., Brechmann, A., and Scheich, H. (2004). Auditory stream segregation relying on timbre involves left auditory cortex. *Neuroreport* 15, 1511–1514. doi: 10.1097/01.wnr.0000132919.12990.34
- Deike, S., Scheich, H., and Brechmann, A. (2010). Active stream segregation specifically involves the left human auditory cortex. *Hear. Res.* 265, 30–37. doi: 10.1016/j.heares.2010.03.005
- Divenyi, P. L., and Danner, W. F. (1977). Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. *Percept. Psychophys.* 21, 125–142. doi: 10.3758/BF03198716
- Dolležal, L.-V., Beutelmann, R., and Klump, G. M. (2012a). Stream segregation in the perception of sinusoidally amplitude-modulated tones. *PLoS ONE* 7:e43615. doi: 10.1371/journal.pone.0043615
- Dolležal, L.-V., Itatani, N., Günther, S., and Klump, G. M. (2012b). Auditory streaming by phase relations between components of harmonic complexes: a comparative study of human subjects and bird forebrain neurons. *Behav. Neurosci.* 126, 797–808. doi: 10.1037/a0030249
- Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., et al. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11073–11078. doi: 10.1073/pnas.0704320104
- Dykstra, A. R., Halgren, E., Thesen, T., Carlson, C. E., Doyle, W., Madsen, J. R., et al. (2011). Widespread brain areas engaged during a classical auditory streaming task revealed by intracranial EEG. *Front. Hum. Neurosci.* 5:74. doi: 10.3389/fnhum.2011.00074
- Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., and Shamma, S. A. (2009). Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron* 61, 317–329. doi: 10.1016/j.neuron.2008.12.005
- Fishman, Y. I., Arezzo, J. C., and Steinschneider, M. (2004). Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *J. Acoust. Soc. Am.* 116, 1656–1670. doi: 10.1121/1.1778903
- Fishman, Y. I., Reser, D. H., Arezzo, J. C., and Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear. Res.* 151, 167–187. doi: 10.1016/S0378-5955(00)00224-0
- Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Essen, D. C. V., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678. doi: 10.1073/pnas.0504136102
- Gerlach, C., Law, I., Gade, A., and Paulson, O. B. (1999). Perceptual differentiation and category effects in normal object recognition A PET study. *Brain* 122, 2159–2170. doi: 10.1093/brain/122.11.2159
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. 1966th Edn. New York: Wiley.
- Grimault, N., Bacon, S. P., and Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *J. Acoust. Soc. Am.* 111, 1340–1348. doi: 10.1121/1.1452740
- Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., and Oxenham, A. J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *J. Neurosci.* 25, 5382–5388. doi: 10.1523/JNEUROSCI.0347-05.2005
- Gutschalk, A., Oxenham, A. J., Micheyl, C., Wilson, E. C., and Melcher, J. R. (2007). Human cortical activity during streaming without spectral cues suggests a general neural substrate for auditory stream segregation. *J. Neurosci.* 27, 13074–13081. doi: 10.1523/JNEUROSCI.12299-07.2007
- Hall, D. A., and Plack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cereb. Cortex* 19, 576–585. doi: 10.1093/cercor/bhn108
- Harinen, K., and Rinne, T. (2013). Activations of human auditory cortex to phonemic and nonphonemic vowels during discrimination and memory tasks. *Neuroimage* 77, 279–287. doi: 10.1016/j.neuroimage.2013.03.064
- Itatani, N., and Klump, G. M. (2009). Auditory streaming of amplitude-modulated sounds in the songbird forebrain. *J. Neurophysiol.* 101, 3212–3225. doi: 10.1152/jn.91333.2008
- Itatani, N., and Klump, G. M. (2011). Neural correlates of auditory streaming of harmonic complex sounds with different phase relations in the songbird forebrain. *J. Neurophysiol.* 105, 188–199. doi: 10.1152/jn.00496.2010
- Kanwal, J. S., Medvedev, A. V., and Micheyl, C. (2003). Neurodynamics for auditory stream segregation: tracking sounds in the mustached bat's natural environment. *Network* 14, 413–435. doi: 10.1088/0954-898X/14/3/303
- Kohlrausch, A., Fassel, R., and Dau, T. (2000). The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. *J. Acoust. Soc. Am.* 108, 723–734. doi: 10.1121/1.429605
- Kondo, H. M., and Kashino, M. (2009). Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *J. Neurosci.* 29, 12695–12701. doi: 10.1523/JNEUROSCI.1549-09.2009
- Micheyl, C., and Oxenham, A. J. (2010). Objective and subjective psychophysical measures of auditory stream integration and segregation. *J. Assoc. Res. Otolaryngol.* 11, 709–724. doi: 10.1007/s10162-010-0227-2
- Micheyl, C., Tian, B., Carlyon, R. P., and Rauschecker, J. P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* 48, 139–148. doi: 10.1016/j.neuron.2005.08.039
- Moore, B. C. J., and Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acust. United Acust.* 88, 320–333.
- Moore, B. C. J., and Gockel, H. E. (2012). Properties of auditory stream formation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 919–931. doi: 10.1098/rstb.2011.0355

- Neff, D. L., Jesteadt, W., and Brown, E. L. (1982). The relation between gap discrimination and auditory stream segregation. *Percept. Psychophys.* 31, 493–501. doi: 10.3758/BF03204859
- Pressnitzer, D., Sayles, M., Micheyl, C., and Winter, I. M. (2008). Perceptual organization of sound begins in the auditory periphery. *Curr. Biol.* 18, 1124–1128. doi: 10.1016/j.cub.2008.06.053
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 676–682. doi: 10.1073/pnas.98.2.676
- Reiterer, S. M., Erb, M., Droll, C. D., Anders, S., Ethofer, T., Grodd, W., et al. (2005). Impact of task difficulty on lateralization of pitch and duration discrimination. *Neuroreport* 16, 239–242. doi: 10.1097/00001756-200502280-00007
- Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., and Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science* 306, 443–447. doi: 10.1126/science.1100301
- Rinne, T., Koistinen, S., Salonen, O., and Alho, K. (2009). Task-dependent activations of human auditory cortex during pitch discrimination and pitch memory tasks. *J. Neurosci.* 29, 13338–13343. doi: 10.1523/JNEUROSCI.3012-09.2009
- Roberts, B., Glasberg, B. R., and Moore, B. C. J. (2002). Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *J. Acoust. Soc. Am.* 112, 2074–2085. doi: 10.1121/1.1508784
- Snyder, J. S., Alain, C., and Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *J. Cogn. Neurosci.* 18, 1–13. doi: 10.1162/089892906775250021
- Szalárdy, O., Bendixen, A., Tóth, D., Denham, S. L., and Winkler, I. (2013). Modulation-frequency acts as a primary cue for auditory stream segregation. *Learn. Percept.* 5, 149–161. doi: 10.1556/LP.5.2013.Suppl2.9
- Thompson, S. K., Carlyon, R. P., and Cusack, R. (2011). An objective measurement of the build-up of auditory streaming and of its modulation by attention. *J. Exp. Psychol. Hum.* 37, 1253–1262. doi: 10.1037/a0021925
- Tregellas, J. R., Davalos, D. B., and Rojas, D. C. (2006). Effect of task difficulty on the functional anatomy of temporal processing. *Neuroimage* 32, 307–315. doi: 10.1016/j.neuroimage.2006.02.036
- Van Noorden, L. P. A. S. (1975). Temporal coherence in the perception of tone sequences. Ph.D. thesis, Eindhoven University of Technology.
- Vliegen, J., Moore, B. C. J., and Oxenham, A. J. (1999). The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *J. Acoust. Soc. Am.* 106, 938–945. doi: 10.1121/1.427140
- Vliegen, J., and Oxenham, A. J. (1999). Sequential stream segregation in the absence of spectral cues. *J. Acoust. Soc. Am.* 105, 339–346. doi: 10.1121/1.424503
- Wilson, E. C., Melcher, J. R., Micheyl, C., Gutschalk, A., and Oxenham, A. J. (2007). Cortical fMRI activation to sequences of tones alternating in frequency: relationship to perceived rate and streaming. *J. Neurophysiol.* 97, 2230–2238. doi: 10.1152/jn.00788.2006

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 19 December 2013; accepted: 03 May 2014; published online: 06 June 2014.
Citation: Dolležal L-V, Brechmann A, Klump GM and Deike S (2014) Evaluating auditory stream segregation of SAM tone sequences by subjective and objective psychoacoustical tasks, and brain activity. *Front. Neurosci.* 8:119. doi: 10.3389/fnins.2014.00119

This article was submitted to Auditory Cognitive Neuroscience, a section of the journal *Frontiers in Neuroscience*.

Copyright © 2014 Dolležal, Brechmann, Klump and Deike. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.