JNeurosci
THE JOURNAL OF NEUROSCIENCE

Research Articles | Behavioral/Cognitive

Perception of tones versus tone streams under random multi-tone masking and its representation in auditory cortex

https://doi.org/10.1523/JNEUROSCI.0416-26.2026

Received: 9 March 2026
Revised: 2 September 2026
Accepted: 21 October 2026

Copyright © 2026 the authors

This Early Release article has been peer reviewed and accepted, but has not been through the composition and copyediting processes.The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Submitted March 9, 2026; Revised September 3, 2026

Perception of tones versus tone streams under random multi-tone masking and its representation in auditory cortex

Frederic Kistenmacher, Laura Doll, Alexander Gutschalk

Department of Neurology, Heidelberg University, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany

Abbreviated title: Perception of streams versus tones

Correspondence to: Alexander.Gutschalk@med.uni-heidelberg.de

Number of pages: 26

Number of Figures: 6

Number of words in Abstract: 250

Introduction: 603

Discussion: 1499

Conflicts of Interest

The authors declare no competing financial interests

JNeurosci Accepted Manuscript

22 **Acknowledgements**

23 Research supported by Deutsche Forschungsgemeinschaft grant DFG 593/5-1 (AG). The data
24 and analysis scripts will be made available on heiDATA, the open research data repository of
25 Heidelberg University (https://heidata.uni-heidelberg.de/dataverse/auditory-cognition-lab).

JNeurosci Accepted Manuscript

Abstract

When repeated tones are presented among random masker tones, their perceptual organization is often variable. In some trials, they cannot be segregated from the masker tones. In other trials, often with a delay, they form coherent streams, segregated from the masker tones. A strong awareness-related negativity is evoked by each segregated tone of the stream in auditory cortex. A similar negative-going response has been observed when subtle, isolated tones are detected versus missed, raising the question of whether target tones that do not form a stream under informational masking are still perceived or not. Here a paradigm is introduced to probe this question while activity in auditory cortex was recorded with magnetoencephalography. Human listeners (13 female, 8 male) were instructed to indicate when they detected an isochronous stream of tones under multi-tone masking while listening to the stimulus. Thereafter, they indicated if a cue tone, presented subsequently to each trial, was comprised in the target and masker combination. Results confirmed that target tones could still be perceived in the subset of trials where they did not form a coherent stream, but the detectability in the cued frequency report was higher for tones forming a stream. A prominent awareness-related negativity was only observed when the tones formed a stream, suggesting that this response reflects the perceptual organization of the tones into streams and beyond the pure perception of the tones themselves. We speculate that the perceptual organization into a stream further enhances the detectability of the tones comprised in the stream.

JNeurosci Accepted Manuscript

Significance Statement

Tone sequences may form streams that segregate from competing aspects of the auditory scene, a phenomenon considered comparable to visual object grouping. In multi-tone masking, hearing out streams of repeated tones seemingly enhances tone detectability by a degree that makes it difficult to judge whether these tones could at all be perceived if they had not formed a coherent stream. We present a paradigm that allows for probing both, the perception of target-tone frequency and the formation of coherent streams in the same trials. Applying this setup in magnetoencephalography showed that the perceptual organization into a stream is more strongly represented in auditory cortex than the ability to correctly detect a tone frequency within a random tone masker.

JNeurosci Accepted Manuscript

Introduction

When sequences of tones are presented among multiple other, random tones, they may sometimes pop out, but be hard to hear at other instances, the latter of which is known as informational masking (Durlach et al., 2003; Kidd et al., 2002). Detected tone sequences in this context come with a prominent negative-going wave in auditory cortex (Gutschalk et al., 2008), evoked by each tone of the detected sequence and referred to as awareness-related negativity (ARN). A similar wave is evoked by single detected tones in the context of near-threshold tones or inattentional deafness (Eklund and Wiens, 2019; Schlossmacher et al., 2021), referred to as auditory awareness negativity (AAN) and conceptualized as a difference wave between aware and unaware trials. Similar activation patterns in auditory cortex were confirmed with fMRI (Dellert et al., 2025; Wiegand et al., 2018) and intracranial EEG (Dykstra et al., 2016). While ARN and AAN are related and may be identical (Dembski et al., 2021), the contexts in which they are observed are quite distinct: The sound intensity of target tones in the multi-tone masker paradigm is typically well above the 50% detection threshold, as is the intensity of the masker tones, but in contrast to inattentional deafness (Schlossmacher et al., 2021), they are still not easily detected in an active, attended detection task. Once the sequence is heard out, listeners usually are aware of all subsequent target tones of the sequence as well (Micheyl et al., 2007). The latter difference is critical, as it suggests that the percept observed is not just the perception of the single tones, but rather the percept of a coherent auditory stream (Bregman, 1990). It is well known that parameters promoting stream segregation reduce informational masking (Kidd et al., 2003, 1994). Moreover, when using random tone sequences as targets instead of an isochronous repetition of the same tone, it was demonstrated that an ARN is only observed when listeners correctly identify the sequence out of the masker tones (Gärtner and Gutschalk, 2021). Accordingly, the different percepts compared in the random multi-tone masker paradigm with sequential targets may be based on two different perceptual organizations, rather than on one between perceived and unperceived single tones: When detected, the target tones are grouped in one stream that is segregated from the masker tones, which form a second, background stream. When masked, all tones are

JNeurosci Accepted Manuscript

perceived as part of the masker. This hypothesis is further supported by the dependence of the ARN amplitude on the inter-stimulus interval (ISI) between the target tones, whereas this ISI has no influence on the amplitude of the evoked response when the tones are undetected and supposedly depend on the timing relation to the masker tones instead (Raudonat et al., 2025).

This hypothesis would thus predict that the target tones are still heard, even when no auditory stream is detected by a listener. Testing this prediction is not trivial, however, since the streaming perception and the detection of the underlying tones need to be acquired in the same trials. Here we present a behavioral setup that probes both: In a stream-report (SR) task, the pop-out of the streams is indicated while listening to the ongoing auditory stimulus (Gutschalk et al., 2008; Micheyl et al., 2007). Thereafter, a frequency cue is provided and listeners indicate in a frequency-report (FR) task whether this tone frequency was part of the tone sequence or not (Demany et al., 2011). Our hypothesis was that the tones can be heard in both cases, but differently. The magnetoencephalogram (MEG) was recorded in parallel, to evaluate how these different percepts are reflected at the level of the auditory cortex.

Methods

Participants

A total of 21 participants (13 female, 8 male) took part in the experiment. They were between 21 and 34 years old, with a mean age of 24.8 years. Two participants were excluded from the study because they did not perform the task correctly (negative d' in SR; details below). Thus, the data of 19 participants were analyzed (11 female, 8 male), aged between 21 and 34 years (mean 24,7 years). All participants reported normal hearing and provided written informed consent. The study was approved by the ethics committee of Heidelberg University, Medical School (S-485/2018).

JNeurosci Accepted Manuscript

Experimental Design and Stimuli

The stimuli were generated with MATLAB (The Mathworks, Natick, MA, USA). All tones had a duration of 100 ms and were gated with a 10-ms Hanning window at the beginning and ending. The sound intensity was approximately 71 dB SPL per tone for all frequencies (measured with continuous tones and corrected for frequency-dependent attenuation of the stimulus device). A sketch of the stimuli is shown in Figure 1.

There were two types of trials: target trials and catch trials. Each trial had a length of 7200 ms. The target trials consisted of 6 isochronous target tones with an interstimulus interval (ISI) of 700 ms. The first target tone began at millisecond 800, the second at millisecond 1600, and so on. The last target tone began at millisecond 4800. All 6 target tones in one trial had the same frequency. Across all trials, there were 18 different target-tone frequencies that were equally distributed on a logarithmic scale between 500 and 2400 Hz.

For each target tone there were 10 masker tones. The masking frequencies were randomly drawn from a range between 100 and 5000 Hz on a logarithmic scale. The onset of the masking tones was jittered in a range of 800 ms around the target tone onsets. Within each target-tone interval, there was a minimum distance of one-twelfth of an octave (i.e. one semitone) between masker tones. This minimum distance was also applied to the masker tones of the following interval, so that there were no masking tones of closer frequency overlapping each other. To reduce energetic masking of the target tones, there was a protected region around the target-tone frequency of a quarter of an octave in both directions, i.e. there were no masking tones in this frequency range of half an octave along the whole trial.

The 6 target-masker intervals were followed by a silent interval of at least 1100 ms. After this silent interval, a cue tone was played, with identical frequency to the target tones but presented without masker tones. Catch trials were structured in the same way as the target trials, but unlike the target trials, they did not comprise a stream of target tones. However, they also had a cue tone at the end and a protected region around the frequency of the cue tone.

JNeurosci Accepted Manuscript

During each trial, listeners were asked to press a button as soon as they detected an isochronous target sequence (SR, stream report). Button presses were acquired up to the onset of the cue tone 1500 ms after the offset of last target tone. After 7200 ms, shortly after the cue tone was played, a response instruction was projected, asking whether the post-trial cue tone was included in the target-masker combination (FR, frequency report). After the question was answered with "yes" or "no", the projection was blanked, and the next trial was played after a delay of 600 ms. For the FR task, listeners were informed that in case they had heard the stream (and supposedly had indicated this in the SR task), they should compare the frequency of the stream's tones with the cue tone. If they did not hear the stream, they should nevertheless indicate if they thought that the cue tone's frequency was part of the masker - target mixture, even as a single tone.

For each of the 18 target-tone frequencies, three different target trials and one catch trial were created. Each of these trials was presented four times in the experiment. The trial order was completely randomized without constraints. Overall, 288 trials were presented to the listeners. 216 (75%) of which were target trials and 72 (25%) were catch trials. The same stimulus set was used for all participants studied.

Participants who did not perform the SR task correctly, i.e. they showed negative d' in SR, were excluded from the analysis, because we expected that the classification in detected- and missed-stream trials could not be used for the substratification required for the FR task. Since the SR task is well established and can usually be performed by most participants, this criterion was not expected to bias the results. In contrast, negative d' in the FR task was tolerated, because it was considered possible that tones were generally not detected in SR miss trials. Excluding negative d' in the FR task would therefore have biased the results towards our hypothesis and was accordingly not implemented.

Data acquisition

JNeurosci Accepted Manuscript

MEG data were acquired with a Neuromag 122-channel whole-head MEG system (Megin, Helsinki, Finland) in a two-layer magnetically shielded room (IMEDCO, Hägendorf, Switzerland). To determine the position of the head in the MEG, four Head Position Indicator (HPI) coils were attached to the mastoids and temples. Using a 3D-space Isotrack II system (Polhemus, Colchester, VT, USA), the position of the HPI coils was digitized relative to three anatomical landmarks (nasion, right and left peri-auricular point), along with 100 additional head-surface points later used as reference for fitting the head model.

The auditory stimuli were played with a custom-made MATLAB script using SoundMexPro (Hörzentrum, Oldenburg, Germany), AD converted with an ADI-8 DS soundcard (RME, Haimhausen, Germany), relayed through PA5 programmable attenuators, and amplified with a HB7 headphone buffer (both from Tucker-Davis Technologies, Alachua, FL, USA). Inside the MEG room, the sounds were presented with ER3 headphones (Etymotic, Elk Grove Village, IL, USA) and conducted to the ears using 97 cm tubes with foam earpieces attached to their ends. Directly before recording was started, the position of the HPI coils in the dewar were measured. The data were recorded continuously with a sampling rate of 1,000 Hz, directly coupled, and with a low-pass filter of 330 Hz.

Classification of trial perception

The events were registered and the MEG data averaged depending on the perception indicated by the listeners: All button presses from the first target tone until the cue tone were considered for SR, but the earliest button presses were observed after the second target tone. For the tone-wise analysis, button presses starting from 150 ms after tone onset were considered to indicate detection of this tone, and the previous tone was also considered detected. This labeling of the two target tones before the button press as hits was based on the consideration that the detection of tone repetition requires at least two tones. The target tones before those two tones were labeled as misses, as were all tones in trials where no button was pressed in SR. To evaluate false alarms for no-signal trials in SR, the onsets where

JNeurosci Accepted Manuscript

188 the target tones would have started were considered, and the number of tones counted as
189 false alarms was determined similar as explained above for hits. All other tones in no-signal
190 trials were considered correct rejections, similar as explained for misses above. For
191 comparison with FR, SR was also evaluated trial wise, i.e. all trials where a button was pressed
192 were considered hits or false alarms, and all trials without a button press were considered
193 misses or correct rejections for signal and no-signal trials, respectively.

194 For the evaluation of FR, all trials with "yes" responses were classified as hits for signal and
195 as false alarms for no-signal trials. All trials with "no" responses were classified as misses for
196 signal trials and as correct rejections for no-signal trials. Moreover, based on the trial-wise
197 evaluation of SR, two groups of trials were separately evaluated in FR: all SR hits and false
198 alarms were used for SR+, whereas all miss trials and correct rejections were used for SR-.
199 The analysis of SR- trials allowed us to evaluate the perception of cue tones in trials where
200 they were not identified as a perceived stream. In contrast, all SR+ trials tested if the target
201 stream reported in SR was of the same frequency as the post-stimulus cue.

202 The detectability (d') was calculated for each category separately. To deal with hit rates and
203 false-alarm rates of zero percent, the log-linear rule was applied (Hautus, 1995), where the hit
204 rates and false-alarm rates inserted were re-calculated with 0.5 added to the numerator and 1
205 added to the denominator.

206 *MEG data analysis*

207 The data were preprocessed and analyzed with MNE-Python version 1.0.3 (Gramfort et al.,
208 2013). First, a high-pass filter of 0.1 Hz and a low-pass filter of 20 Hz (zero phase shift, finite-
209 impulse-response filters with 6 dB roll off) were applied. Filters were used to improve the signal-
210 to-noise ratio and the low-pass filter cutoff was previously evaluated to cause no relevant
211 distortion to the ARN time course. The raw waveforms of each participant were then visualized
212 and bad channels were manually identified and interpolated. Signal-space projection was used
213 to reduce low-frequency environmental noise, mostly caused by nearby streetcars (Uusitalo
214 and Ilmoniemi, 1997). Cardiac and ocular artifacts were estimated using independent

JNeurosci Accepted Manuscript

component analysis (Hyvarinen, 1999). If components were removed, the sensor signals were subsequently reconstructed. Then the epochs were averaged according to the behavioral categories outlined above into epochs starting 100 ms before and ending 1000 ms after tone onset. Epochs with artifacts, identified by high amplitude, were repaired or rejected using the autoreject module from Python (Jas et al., 2017). The average rejection threshold was 2800 fT/cm (SD: 2300 fT/cm). On average, 6.5 % of the epochs were rejected.

The source analysis was based on a minimum-norm estimate (MNE) with depth weighting (Lin et al., 2006), calculated on an average cortical source space based on the FS-average brain from FreeSurfer (Fischl et al., 1999). To this end, the head surface of FS average was scaled in X, Y, and Z dimension to match the coordinates of the head digitization. The individually rescaled head and segmented brain were then used for calculating the boundary-element head model, and the individually scaled and segmented brain was used to estimate the individual source space. The noise covariance matrix was calculated for 100 ms baseline intervals starting 100 ms before the onset of each tone across all trials. The MEG data were then evaluated with a region of interest (ROI) in the left and right auditory cortex. The ROI covers the auditory cortex including Heschl's gyrus and planum temporale (Glasser et al., 2016). The source time courses in this ROI were first averaged across all dipoles at the single listener level for the left and right auditory cortex. The amplitude of the ARN was then measured in the time interval 75–175 ms, and additionally in the interval 175–275 ms with reference to previous studies (Gärtner and Gutschalk, 2021; Gutschalk et al., 2008; Königs and Gutschalk, 2012). In these studies, the earlier time interval also covered the N1 evoked by targets without masking and the ARN showed similar behavior on lateralization as the N1. In the later time interval, the ARN - as defined by its presence for detected streams - still continues, but this interval does not usually include the N1 evoked by unmasked trials, but rather some less well characterized later negative-going components as well as the P2 and was therefore separated from the earlier interval.

JNeurosci Accepted Manuscript

Statistical Analysis

The sample size was set to 20 participants based on previous MEG studies that had reliably found the ARN with a sample size of 12–20 participants (Dykstra and Gutschalk, 2015; Gutschalk et al., 2008; Königs and Gutschalk, 2012), when the trials were evaluated tone wise. The study was not powered to detect differences between detected and undetected tones in auditory cortex, when the tones were not perceived in SR. The best estimate we had for the expected effect size here was based on a previous study with post-stimulus frequency cues (Gärtner and Gutschalk, 2021), which yielded an estimate of N=328 for the right and N=571 for the left auditory cortex to show a difference between hit and miss trials (Cohen's d=0.117 on the left and d=0.155 on the right). Because we did not have resources to study a sample coming close to these numbers, we decided to evaluate this comparison on an exploratory basis, only. In contrast, the behavioral effect was already found robustly in a sample of 20 participants in the latter study.

The statistical analysis of the behavioral data and the ROI analysis was performed with Python 3.10.9, using the SciPy library and the Pingouin package. ROI analyses were evaluated with an ANOVA for repeated measures including the factors detection (hit, miss) and hemisphere (left, right). In an additional analysis detailed in the supplement, the six subsequent target tones were separately averaged and modeled in the ANOVA with the additional factor tone position.

The main hypothesis for the behavioral data was that FR hits that were not previously heard as a stream (SR-) could still be perceived. If this was the case, we expected two parallel effects: (1) that the hit rate was higher than the false alarm rate, and (2) that the d' was consequently larger than and significantly different from zero. Both were tested with planned t-tests, (1) with a paired-sample, two-tailed t-tests, (2) with a single-sample, two-tailed t-test. The second hypothesis was that the perception of a stream increased the d' for FR compared to trials not perceived as a stream. This hypothesis was also tested with a planned, paired-sample, two-tailed t-test. Further comparisons are purely descriptive. No correction for multiple comparison was applied for the t-tests reported in this manuscript.

Neurosci Accepted Manuscript

Signal-Detection-Theory Models

The behavioral results were used to predict the signal strength using signal-detection theory (Green and Swets, 1966). The basic assumption is that signal strengths in response to both signal and noise trials are normally distributed, but the mean of the signal distribution is higher compared to the noise distribution. Hits and misses (and correspondingly false alarms and correct rejections) are distinguished based on whether the signal is larger or smaller than the response criterion. This criterion is – usually subconsciously – set by the participant.

For our experiment, we compared two different models with the acquired data: Model 1 assumed that there are two criteria for the two tasks: In trials with sufficient signal strength to cross the higher, more conservative, criterion 1, the target was already detected in the SR task. If signal strength was lower, but still strong enough to cross the lower criterion 2, the target was perceived and correctly detected in the FR task, only. All other trials were not detected in either task. The mean of the noise distribution μ_{noise} was arbitrarily set to 0 and the standard deviation σ_{noise} to 1, since only the relation between noise and signal distribution matters. The distance between both distributions is $\mu_{signal} - \mu_{noise}$ defined as d', here we used the d' of the FR task for the complete data set (i.e. without substratification). The standard deviation σ_{signal} can be derived from the empirical value of $(\mu_{signal} - \mu_{noise})/(\sigma_{ssignal} - \sigma_{snoise}) = 4$. With the resulting signal distribution, the criteria were calculated to match the observed detection rates: The area under the signal distribution above criterion 1 corresponds to the hit rate FR hits (SR+). The area under the curve between criterion 2 (which is lower, i.e. more liberal than criterion 1) and criterion 1 corresponds to the hit rate FR hits (SR-). Finally, the probability of FR misses (SR-) is reflected in the area under the curve below criterion 2. The corresponding amplitudes are the centroids of those intervals. Since those values are in arbitrary units, we divided both the predicted and the measured amplitudes by the respective FR hits (SR-) amplitudes for normalization.

JNeurosci Accepted Manuscript

Model 2 assumed that SR+ and SR- trials are reflected by two separate signal distributions when modeling the FR task. Accordingly, the mean of the right gaussian is based on the d' of FR for SR+ trials and the left on the d' for FR for SR- trials. The standard deviation is empirically estimated with the formula provided above. The response criterion is then based on the false-alarm rate across all FR trials.

Results

Behavioral data

The trial-wise behavioral results are summarized in Figure 2, showing hit rates and false-alarm rates in panel A, sensitivity (d') in panel B and response criteria (c) in panel C. The average trial-wise hit rate in SR was 49%, which is almost ideal to separate two subgroups for the analysis of FR. The trial-wise hit rate in FR was on average 77% and higher compared to SR (t_{18}=-6.606, p<0.001, d=1.516, η^2_p =0.71), indicating that some target trials were rated as detected for which an isochronous target stream was not previously identified in SR. Since the false-alarm rate was also higher in FR compared to SR (t_{18}=-3.136, p=0.006, d=0.719, η^2_p =0.353), however, the d' was not significantly different between the two tasks (t_{18}=-1.653, p=0.116, d=0.379, η^2_p =0.132).

Next, we separately evaluated FR only for trials in which listeners had not indicated hearing an isochronous target stream in SR (SR-), to test our main hypothesis that target-tone and target-stream perception can be distinguished. Results showed an average hit rate of 58% for this case with a significantly smaller false-alarm rate of 22% (t_{18}=8.867, p<0.001, d=2.034, η^2_p =0.814). This yielded a d' of 1.2, which was significantly different from zero across listeners (no negative values observed; t_{18}=7.935, p<0.001, d=1.82, η^2_p =0.778). The SR- subset of the FR task thus confirms that target tones could be perceived within the multi-tone masker, even when listeners did not indicate perception of the isochronous tone stream before. The SR- subsample of the FR was then compared to the SR+ subsample, i.e. all trials in which listeners

JNeurosci Accepted Manuscript

had previously indicated to perceive an isochronous tone stream. Our second hypothesis was that the d' in FR would be higher for SR+ compared to SR-, because the perception of the tones in a separate stream made them more salient. The average d' for FR detection was 2.3 for SR+, compared to 1.2 for SR-, and the difference was significant (t_{18}=-3.355, p=0.004, d=0.77, η^2_p =0.385). A problem for calculating d' for SR+ is that the number of no-signal trials was small, since only false-alarm trials of SR were considered for their estimate. Three listeners did not have false alarms in SR and their false-alarm rate was therefore also set to zero in the SR+ subsample of FR shown in Figure 2. Since this could lead to a bias towards our hypothesis, we tested again with these three listeners excluded (Fig. S1). Results showed that d' was still higher for SR+ in the remaining 16 subjects (t_{15}=-2.778, p=0.014, d=0.695, η^2_p =0.34). A number of participants showed a very high false-alarm-rate in FR (SR+), because they had more false alarms per trial in both tasks than correct rejections in FR after a false alarm in SR. Note that these participants had all been included in this comparison. The false alarm rate was numerically somewhat larger in the SR+ compared to the SR- subset, with a trend for statistical significance (t_{18}=-1.813, p=0.087, d=0.416, η^2_p =0.154). When the behavioral results were evaluated in three intervals of equal duration along the experiment, all hit and false-alarm rates remained quite stable (Fig. S2), suggesting that the false alarm rates are reliable despite the small number of no-signal trials in the SR+ subset. If the common false-alarm rate, based on all no-signal trials, was used instead, the d' for FR in the SR hits subsample would be estimated to be somewhat higher (d'=2.8).

It was previously found that the detectability of target streams decreases for frequencies beyond 1000 Hz (Gutschalk et al., 2008). The present study used constant sound intensity across frequencies, whereas the previous study used constant sensation level. Despite the resulting higher sound intensity for high-frequency tones in the present study, the decreasing hit rate in the SR task for high frequencies was reproduced (Fig. 3A, C). As a result, more SR- tones are available in the high- compared to the low-frequency range (Fig. 3A). The hit rate in the FR task of the SR- subset (Fig. 3C) did not show a similarly strong reduction for high frequencies as observed for the SR task.

JNeurosci Accepted Manuscript

MEG data

The MEG source waveforms calculated for the ROIs in left and right auditory cortex are shown in Fig. 4 for the tone-wise analysis based on the SR task; whole-brain source analysis maps are shown in Fig. S3A. The corresponding tone-wise hit rate in SR was 32% (SD = 0.16), the tone-wise false-alarm rate was 5% (SD = 0.08). The results replicate previous findings of a strong ARN for all tones grouped into a stream, which was significantly stronger than the response observed for tones not grouped into a stream in the time interval 75–175 ms (repeated measures ANOVA with factors detection (hit, miss) and hemisphere (left, right), detection: $F_{1,18}$=9.809, p=0.006, η^2_p=0.353). The overall auditory cortex activity was significantly stronger on the right (hemisphere: $F_{1,18}$=5.298, p=0.034, η^2_p=0.227), but there was no significant interaction of detection with hemisphere ($F_{1,18}$=0.160, p=0.694, η^2_p=0.009) in this analysis. Similar results were observed in the subsequent time window 175–275 ms (detection: $F_{1,18}$=6.622, p=0.019, η^2_p=0.269; hemisphere: $F_{1,18}$=3.890, p=0.064, η^2_p=0.178; hemisphere x detection: $F_{1,18}$=1.270, p=0.275, η^2_p=0.066).

The trial-wise analysis of the MEG data is shown in Fig. 5; here, all 6 tones of a trial were averaged based on the behavioral response in the FR task, for subgroups based on the SR task. Miss trials in FR for the SR+ subsample were not evaluated because of too small number. Moreover, datasets where less than five hit or miss trials occurred in the SR- subsample were excluded from the MEG analysis, because such low numbers of averages do not provide a sufficient signal-to-noise ratio for the analysis. One participant was thus excluded from the analysis because of low hit rates and another one because of low miss rates in the SR- subsample, reducing the number of evaluated data sets to 17 for this part of the analysis (cf. Fig 3B).

The comparison FR hits (SR+) versus FR misses (SR-) revealed stronger MEG activity in the auditory cortex, similar to the tone-wise analysis in the 75–175 ms time window (repeated

JNeurosci Accepted Manuscript

measures ANOVA with factors detection (hit, miss) and hemisphere (left, right), detection: $F_{1,16}$=12.559, p=0.003, η^2_p=0.440) with stronger overall activity in the right (hemisphere: $F_{1,16}$=7.843, p=0.013, η^2_p=0.329) and a trend for a hemisphere x detection interaction ($F_{1,16}$=3.505, p=0.080, η^2_p=0.180). The detection effect was also significant in the later time window (detection: $F_{1,16}$=6.243, p=0.024, η^2_p=0.281), but there were no significant hemisphere effects (hemisphere: $F_{1,16}$=0.720, p=0.409, η^2_p=0.043; detection x hemisphere: $F_{1,16}$=1.770, p=0.202, η^2_p=0.100).

Note that the experiment was not powered for separating hits and misses in the SR- subsample in MEG. Therefore, this part of the statistical analysis is purely exploratory: The comparison of FR hits and misses from the SR- subsample was done with a repeated measures ANOVA with factors hemisphere (left, right) and detection (hit, miss); post-hoc *t*-tests are not corrected for multiple comparison. The only significant effect was a main effect of hemisphere in the 75–175 ms time interval ($F_{1,16}$=10.337, p=0.005, η^2_p=0.392). There was no significant effect of detection in either time interval (75-175 ms: $F_{1,16}$=1.763, p=0.203, η^2_p=0.099; 175-275 ms: $F_{1,16}$=1.459, p=0.245, η^2_p=0.084). However, there was a trend for a detection x hemisphere interaction in the later time interval (175-275 ms: $F_{1,16}$=4.191, p=0.057, η^2_p=0.208; 75-175 ms: $F_{1,16}$=2.698, p=0.120, η^2_p=0.144). We therefore tested the detection effect separately in the right and left auditory cortex, which revealed in right auditory cortex a significant difference between hits and misses in the late time window (175-275ms: t_{16}=2.266, p=0.038, *d*=0.55, η^2_p=0.243) and a trend in the early time window (75-175ms: t_{16}=2.046, p=0.058, *d*=0.496, η^2_p=0.207). No indication of a detection effect was observed in the left auditory cortex (75-175ms: t_{16}=-0.300, p=0.768, *d*=-0.073, η^2_p=0.006; 175-275ms: t=-1.007, p=0.329, *d*=-0.244, η^2_p=0.06). We further explored if the trend observed in the average across tones 1 - 6 was more prominent for specific tone positions. The results of this analysis are detailed in supplemental Fig. S5: While the trend for stronger responses for FR hits versus misses in SR- trials in right auditory cortex was confirmed, a significant dependence on tone position was not observed, in particular there was no indication for a delayed buildup of a negative-going wave.

JNeurosci Accepted Manuscript

In summary, these data cannot confirm that FR hit and miss trials in the SR- subset are differently represented in the auditory cortex, as was already expected by previous power estimates. The exploratory analysis suggests that a difference might exist in the right auditory cortex and that the effect could be somewhat stronger (d=0.496) than estimated from a previous experiment (d=0.155) with a different masker structure (Gärtner and Gutschalk, 2021). Note that target-tone frequency was not controlled between conditions (Fig 3A) in the MEG analysis, because this would have further reduced the signal-to-noise ratio. The N1 and ARN amplitudes have been previously shown to decrease with tone frequency (Gutschalk et al., 2008), but the ARN was still observed when hits and misses were compared per frequency. In the present data, the frequency imbalance between SR+ and SR- could bias SR+ to show somewhat higher amplitudes. For the comparison of SR- hit and miss trials in the FR task, the frequency imbalance is only modest.

Model comparison

Our hypothesis for the comparison of SR+ and SR- trials is that they represent two qualitatively different percepts, in which the streaming perception places the tone frequency in a completely different context (target stream in SR+ versus masker in SR-). Alternatively, the tone perception could be the continuum of a single perceptual quality and neural representation, where the impression of a segregated stream emerges in trials where the internal signal is high and the tones are perceived more clearly.

Figure 6 shows two models based on signal-detection theory with the behavioral data as input, which are based on the latter consideration of a single perceptual quality. Model 1 (Fig. 6A) is based on the assumption that the SR+ subsample simply selects trials with a stronger internal representation by applying a more conservative response criterion, such as in multi-level rating tasks (Doll et al., 2024). A single gaussian distribution is considered to comprise both FR subsamples (SR+ and SR-), and the criteria are estimated based on the hit rates, such that FR hits (SR+) are to the right of the higher criterion, FR hits (SR-) lie between the higher and

JNeurosci Accepted Manuscript

the lower criterion, and FR misses (SR-) are to the left of the lower criterion. One problem for this model is that the conservative criterion for FR hits (SR+) does not match with the liberal criterion that we found on average in the experimental data (Fig. 2C).

Model 2 (Fig. 6C) uses separate distributions for SR+ and SR- trials, while the response criterion is based on the false alarm rate observed in the FR task across all trials. This configuration would for example be expected when attention shifts the d' of an internal representation (Carrasco et al., 2004). Provided the similar false alarm rates in both subsamples, this model matches the empirical c much better (note that c is relative to the intersection of the noise and signal gaussian).

Next, the model predictions for the internal signal (Fig. 6B, D) can be compared to the amplitudes of the evoked response in auditory cortex. Since the absolute baseline in this model is arbitrary, it is more meaningful to compare differences. We therefore calculated the difference of FR hit minus FR miss (SR-) separately for the SR+ and SR- subsamples and then calculated the ratio of SR+/SR- to be independent of absolute scales. This ratio indicates how much larger the enhancement is on average for SR+ hits compared to SR- hits. For the models, the ratio amounted to 2.2 in model 1 and to 1.4 in model 2. In the MEG data, the ratio amounted to 3.6 in right auditory cortex and to -13.8 in left auditory cortex (negative because hit smaller than miss in the SR- subsample). Across both hemispheres, the ratio amounted to 29.7 in MEG. Thus, even in the case of right auditory cortex, where a trend for enhanced negative-going activity is observed for SR- hits, the response to SR+ trials perceived as stream is 3.6 times stronger, whereas the models would only suggest a range of 1.4-2.2, so that the models can only explain part of the strong asymmetry between SR+ and SR- hit trials.

Discussion

The behavioral results confirm our hypothesis that target tones presented in a jittered multi-tone masker are still perceived, even when they do not form a stream that segregates the

JNeurosci Accepted Manuscript

tones from the masker-tone cloud. The MEG data further demonstrate that the ARN evoked by tones perceived in a stream is much more prominent than a potential negative-going response for detected tones when no stream perception was indicated. As a consequence, we argue that the paradigm reflects the perception of auditory streams in addition to and more strongly than the perception of the tones themselves.

Attention within a continuous representation

Doll et al (2024) suggested that the amplitude variation of the AAN evoked by single tones in noise can be modeled by the range of the decision criteria chosen in a rating task. When streaming perception was similarly used to indicate a more conservative response criterion in model 1 (Fig. 6A), two aspects of the data were not explained well. First, the observed response criterion was not as conservative as suggested by the model. Second, it could not fully explain the strong response enhancement observed for SR+ over SR- in MEG. The criterion variability was previously explained by fluctuating arousal or sustained attention. When selective, top-down attention is considered to potentially act on top of such fluctuation, a shift of d' would be expected (Carrasco et al., 2004), but still within the same neural representation. This was implemented in model 2 (Fig. 6C), which complied with the observed decision criteria, but did not explain the neural asymmetry observed for SR+ over SR- trials, either. Thus, while attention that amplifies an otherwise identical representation might explain part of the enhancement observed for SR+ and possibly the higher d' observed in this subset, it does not explain the much stronger ARN for SR+ in the MEG data.

Awareness and perceptual organization

We therefore argue that the ARN does not only reflect the perception of the isolated or attended tones, but of tones grouped into distinct auditory streams (Bregman, 1990), which separates

JNeurosci Accepted Manuscript

them from the other, random masker tones. We further argue that this organization into streams interacts with the representation of time intervals within each stream. It has been shown that the ARN evoked by detected streams increases with the ISI (Raudonat et al., 2025), but remains independent of the target ISI as long as it remains undetected. The response evoked by undetected target tones under multi-tone masking was previously shown to be similar to the response evoked by masker tones. Since at least some of the masker tones are undoubtedly perceived, it is plausible that target tones that are not perceived as part of the target stream are simply grouped into the masker and can be detected within the latter, as well.

This raises the question if target tones that are not segregated from the masker but still perceived should not evoke an AAN, as is suggested by the SDT model in Fig. 6, and if the AAN and ARN are the same or different phenomena at all (Dembski et al., 2021). First, while the AAN expected in FR hit (SR-) trials could not be confirmed in the present data, the results still suggest that it may be present, albeit small and only in right auditory cortex. Second, while the AAN has so far been considered to reflect the perception of single tones, the ISI between near-threshold tones was in a range of at least a second (Eklund et al., 2020; Eklund and Wiens, 2019). The situation is less clear in inattentional deafness paradigms (Schlossmacher et al., 2021); while the ISI between distractors and targets was short, the two were easily segregated and the ISI between subsequent targets was also longer than a second. If the AAN and ARN are the same, this could explain why the AAN in SR- trials was only small given the small ISI relative to other (unsegregated) masker tones. Third, considering the information load inherent to the multi-tone masker, performance in the FR task is not only limited by perception and masking, but likely also by memory (Block, 2007). Segregated tones are easy to remember, and their report is likely more reliable. It is hard to estimate the percentage in which unsegregated target tones, and similarly masker tones, were perceived as part of the masker, but their frequency was not remembered by the time the post-hoc cue was presented. In the most extreme case, all tones were generally perceived, and the report was limited to those that somehow stand out and maybe evoke a slightly stronger N1 or AAN. Such a scenario would make the FR misses (SR-) an unreliable estimate of unperceived tones, because they were

JNeurosci Accepted Manuscript

then perceived but not remembered. Conversely, if only a fraction of the masker tones was perceived, which is the scenario originally assumed, then it was also unlikely that all six SR- target tones are perceived in hit trials. It might suffice if one or two of six tones were perceived, whereas four of six target tones were on average perceived in SR+ trials. This could reduce the evoked response for SR- trials to a quarter of SR+ trials in the most extreme case.

Representation of auditory features and streams

It has been argued that the visual awareness negativity (VAN) represents the perception of lower-level features (Jimenez et al., 2018; Koivisto et al., 2017; Wiens et al., 2023). On a higher level, the conscious perception of faces covaries with the N170 in fusiform gyrus and the VAN in occipital cortex (Dellert et al., 2021; Shafto and Pitts, 2015). The VAN follows the N170 in latency, despite its supposedly lower level, supporting a potential role as a marker of recurrent-processing (Supèr et al., 2001).

An example for a lower-level auditory representation is sound lateralization for the ARN, N1 (Königs and Gutschalk, 2012) and AAN (Eklund et al., 2021). An AAN example of higher-level representation is the perception of sine-wave speech (Zhu et al., 2024). For the latter, source analysis is not available, and the sources of non-ambiguous vowel-specific responses are found close to those of the ARN in MEG (Gutschalk and Uppenkamp, 2011). Because of this lack of an anatomical dissociation of lower- and higher-level representation in auditory cortex, we can currently not evaluate if the AAN/ARN specifically represents the higher-level percept or the feedback to the lower-level representation.

For sequential grouping, (Bregman, 1990), which we argue is related to the ARN in the present setup, an anatomically distinct neural representation is also unlikely. In animal models, frequency-selective adaptation in early auditory cortex plays a role (Fishman et al., 2001; Micheyl et al., 2005), but does not directly reflect perception (Banno et al., 2023). The activity observed in the latter studies in monkey is generated in middle layers and more likely related to the P1 in human (Gutschalk et al., 2005), whereas the ARN/AAN is generated in superficial

JNeurosci Accepted Manuscript

layers (Dykstra et al., 2026). We suggest that the sequence-specific modulation of the ARN may be important for the conscious perception of perceived streams, in parallel to the content of the single sound elements.

Interaction of attention and perceptual organization

There is good evidence for a role of top-down attention for the ARN in tone streams under multi-tone masking (Elhilali et al., 2009; Gutschalk et al., 2008), as well as the AAN for single tones or words in noise (Doll et al., 2024; Schlossmacher et al., 2021). A potential source of this interaction lies in the insular and anterior-medial cingulate cortex (Dellert et al., 2025; Doll et al., 2025; Sadaghiani et al., 2009). We expect that attention is required in these ambiguous scenes, and that it may selectively enhance the representation of the target stream, as suggested in the biased-competition framework (Desimone and Duncan, 1995). In the present case, where the frequency of the target stream is unknown, directing attention to a random frequency is unlikely a successful strategy (Micheyl et al., 2007). Conversely, selective listening for an isochronous tone repetition would require that the stream is segregated from the multi-tone masker before it is attended. Attention might then further enhance the whole stimulus including the spectral representation, in analogy to vision (Schoenfeld et al., 2014; Shinn-Cunningham, 2008). Based on these considerations, we think that a streaming-specific mechanism is likely involved in the generation of the ARN, but based on the present data, a further distinction of attention and segregation mechanisms for the ARN is not possible.

Limitations

There are two limitations that should be considered when interpreting these results. First, the study was not powered to detect a neural representation of detected tones that did not form an auditory stream (FR hits versus misses, SR-). Therefore, while the present results leave the possibility that a neural correlate of these FR detections might be observed in the right

JNeurosci Accepted Manuscript

auditory cortex, this hypothesis requires confirmation including control of tone-frequency effects. Second, the relative roles of stream formation and attention for the generation of the ARN cannot be dissociated based on the present data.

References

Banno, T., Shirley, H., Fishman, Y.I., Cohen, Y.E., 2023. Changes in neural readout of response magnitude during auditory streaming do not correlate with behavioral choice in the auditory cortex. Cell Reports 42, 113493. https://doi.org/10.1016/j.celrep.2023.113493

Block, N., 2007. Consciousness, accessibility, and the mesh between psychology and neuroscience. Behav Brain Sci 30, 481–499. https://doi.org/10.1017/S0140525X07002786

Bregman, A.S., 1990. Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press, Cambridge, USA.

Carrasco, M., Ling, S., Read, S., 2004. Attention alters appearance. Nat Neurosci 7, 308–313. https://doi.org/10.1038/nn1194

Dellert, T., Balster, H., Schlossmacher, I., Bruchmann, M., Moeck, R., Straube, T., 2025. Neural correlates of consciousness in an auditory no-report fMRI study. Current Biology 35, 5721-5733.e3. https://doi.org/10.1016/j.cub.2025.10.026

Dellert, T., Müller-Bardorff, M., Schlossmacher, I., Pitts, M., Hofmann, D., Bruchmann, M., Straube, T., 2021. Dissociating the Neural Correlates of Consciousness and Task Relevance in Face Perception Using Simultaneous EEG-fMRI. J. Neurosci. 41, 7864–7875. https://doi.org/10.1523/JNEUROSCI.2799-20.2021

Demany, L., Semal, C., Pressnitzer, D., 2011. Implicit versus explicit frequency comparisons: Two mechanisms of auditory change detection. Journal of Experimental Psychology: Human Perception and Performance 37, 597–605. https://doi.org/10.1037/a0020368

Dembski, C., Koch, C., Pitts, M., 2021. Perceptual awareness negativity: a physiological correlate of sensory consciousness. Trends in Cognitive Sciences 25, 660–670. https://doi.org/10.1016/j.tics.2021.05.009

Desimone, R., Duncan, J., 1995. Neural Mechanisms of Selective Visual Attention. Annu. Rev. Neurosci. 18, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Doll, L., Dykstra, A.R., Gutschalk, A., 2024. Perceptual awareness of near-threshold tones scales gradually with auditory cortex activity and pupil dilation. iScience 27, 110530.

JNeurosci Accepted Manuscript

https://doi.org/10.1016/j.isci.2024.110530

Doll, L., Heiland, S., Gutschalk, A., 2025. A Role of Pupil-linked Arousal, Cingulo-insular Cortex, and Intralaminar Thalamus for Auditory Near-threshold Perception. Journal of Cognitive Neuroscience 1–25. https://doi.org/10.1162/jocn_a_02324

Durlach, N.I., Mason, C.R., Kidd, G., Arbogast, T.L., Colburn, H.S., Shinn-Cunningham, B.G., 2003. Note on informational masking (L). J. Acoust. Soc. Am. 113, 2984. https://doi.org/10.1121/1.1570435

Dykstra, A.R., Gutschalk, A., 2015. Does the mismatch negativity operate on a consciously accessible memory trace? Sci. Adv. 1, e1500677. https://doi.org/10.1126/sciadv.1500677

Dykstra, A.R., Halgren, E., Gutschalk, A., Eskandar, E.N., Cash, S.S., 2016. Neural Correlates of Auditory Perceptual Awareness and Release from Informational Masking Recorded Directly from Human Cortex: A Case Study. Front. Neurosci. 10. https://doi.org/10.3389/fnins.2016.00472

Dykstra, A.R., Zhu, Y., Fernandez Pujol, C., Zhou, D.W., Jones, S.R., Marvan, T., Bonaiuto, J.J., 2026. Testing circuit-level theories of consciousness in humans. Trends in Cognitive Sciences 30, 226–238. https://doi.org/10.1016/j.tics.2025.08.012

Eklund, R., Gerdfeldter, B., Wiens, S., 2021. The early but not the late neural correlate of auditory awareness reflects lateralized experiences. Neuropsychologia 158, 107910. https://doi.org/10.1016/j.neuropsychologia.2021.107910

Eklund, R., Gerdfeldter, B., Wiens, S., 2020. Is auditory awareness negativity confounded by performance? Consciousness and Cognition 83, 102954. https://doi.org/10.1016/j.concog.2020.102954

Eklund, R., Wiens, S., 2019. Auditory awareness negativity is an electrophysiological correlate of awareness in an auditory threshold task. Consciousness and Cognition 71, 70–78. https://doi.org/10.1016/j.concog.2019.03.008

Elhilali, M., Xiang, J., Shamma, S.A., Simon, J.Z., 2009. Interaction between Attention and Bottom-Up Saliency Mediates the Representation of Foreground and Background in an Auditory Scene. PLoS Biol 7, e1000129. https://doi.org/10.1371/journal.pbio.1000129

Fischl, B., Sereno, M.I., Tootell, R.B.H., Dale, A.M., 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8, 272–284. https://doi.org/10.1002/(SICI)1097-0193(1999)8:4<272::AID-HBM10>3.0.CO;2-4

Fishman, Y.I., Reser, D.H., Arezzo, J.C., Steinschneider, M., 2001. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hearing Research 151, 167–187. https://doi.org/10.1016/S0378-5955(00)00224-0

JNeurosci Accepted Manuscript

627 Gärtner, K., Gutschalk, A., 2021. Auditory cortex activity related to perceptual awareness
628 versus masking of tone sequences. NeuroImage 228, 117681.
629 https://doi.org/10.1016/j.neuroimage.2020.117681

630 Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil,
631 K., Andersson, J., Beckmann, C.F., Jenkinson, M., Smith, S.M., Van Essen, D.C., 2016. A
632 multi-modal parcellation of human cerebral cortex. Nature 536, 171–178.
633 https://doi.org/10.1038/nature18933

634 Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, Brodbeck, C., Goj, R.,
635 Jas, M., Brooks, T., Parkkonen, L., Hämäläinen, M., 2013. MEG and EEG data analysis with
636 MNE-Python. Front. Neurosci. 7. https://doi.org/10.3389/fnins.2013.00267

637 Green, D.M., Swets, J.A., 1966. Signal Detection Theory and Psychophysics. John Wiley,
638 Oxford.

639 Gutschalk, A., Micheyl, C., Melcher, J.R., Rupp, A., Scherg, M., Oxenham, A.J., 2005.
640 Neuromagnetic Correlates of Streaming in Human Auditory Cortex. J. Neurosci. 25, 5382–
641 5388. https://doi.org/10.1523/JNEUROSCI.0347-05.2005

642 Gutschalk, A., Micheyl, C., Oxenham, A.J., 2008. Neural Correlates of Auditory Perceptual
643 Awareness under Informational Masking. PLoS Biol 6, e138.
644 https://doi.org/10.1371/journal.pbio.0060138

645 Gutschalk, A., Uppenkamp, S., 2011. Sustained responses for pitch and vowels map to
646 similar sites in human auditory cortex. NeuroImage 56, 1578–1587.
647 https://doi.org/10.1016/j.neuroimage.2011.02.026

648 Hautus, M.J., 1995. Corrections for extreme proportions and their biasing effects on
649 estimated values ofd′. Behavior Research Methods, Instruments, & Computers 27, 46–51.
650 https://doi.org/10.3758/BF03203619

651 Hyvarinen, A., 1999. Fast and robust fixed-point algorithms for independent component
652 analysis. IEEE Trans. Neural Netw. 10, 626–634. https://doi.org/10.1109/72.761722

653 Jas, M., Engemann, D.A., Bekhti, Y., Raimondo, F., Gramfort, A., 2017. Autoreject:
654 Automated artifact rejection for MEG and EEG data. NeuroImage 159, 417–429.
655 https://doi.org/10.1016/j.neuroimage.2017.06.030

656 Jimenez, M., Grassini, S., Montoro, P.R., Luna, D., Koivisto, M., 2018. Neural correlates of
657 visual awareness at stimulus low vs. high-levels of processing. Neuropsychologia 121, 144–
658 152. https://doi.org/10.1016/j.neuropsychologia.2018.11.001

659 Kidd, G., Mason, C.R., Arbogast, T.L., 2002. Similarity, uncertainty, and masking in the
660 identification of nonspeech auditory patterns. The Journal of the Acoustical Society of

JNeurosci Accepted Manuscript

America 111, 1367–1376. https://doi.org/10.1121/1.1448342

Kidd, G., Mason, C.R., Deliwala, P.S., Woods, W.S., Colburn, H.S., 1994. Reducing informational masking by sound segregation. The Journal of the Acoustical Society of America 95, 3475–3480. https://doi.org/10.1121/1.410023

Kidd, G., Mason, C.R., Richards, V.M., 2003. Multiple bursts, multiple looks, and stream coherence in the release from informational masking. J. Acoust. Soc. Am. 114. https://doi.org/10.1121/1.1621864

Koivisto, M., Grassini, S., Salminen-Vaparanta, N., Revonsuo, A., 2017. Different Electrophysiological Correlates of Visual Awareness for Detection and Identification. Journal of Cognitive Neuroscience 29, 1621–1631. https://doi.org/10.1162/jocn_a_01149

Königs, L., Gutschalk, A., 2012. Functional lateralization in auditory cortex under informational masking and in silence: Spatial informational masking. European Journal of Neuroscience 36, 3283–3290. https://doi.org/10.1111/j.1460-9568.2012.08240.x

Lin, F.-H., Witzel, T., Ahlfors, S.P., Stufflebeam, S.M., Belliveau, J.W., Hämäläinen, M.S., 2006. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. NeuroImage 31, 160–171. https://doi.org/10.1016/j.neuroimage.2005.11.054

Micheyl, C., Shamma, S.A., Oxenham, A.J., 2007. Hearing Out Repeating Elements in Randomly Varying Multitone Sequences: A Case of Streaming?, in: Kollmeier, B., Klump, G., Hohmann, V., Langemann, U., Mauermann, M., Uppenkamp, S., Verhey, J. (Eds.), Hearing – From Sensory Processing to Perception. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 267–274. https://doi.org/10.1007/978-3-540-73009-5_29

Micheyl, C., Tian, B., Carlyon, R.P., Rauschecker, J.P., 2005. Perceptual Organization of Tone Sequences in the Auditory Cortex of Awake Macaques. Neuron 48, 139–148. https://doi.org/10.1016/j.neuron.2005.08.039

Raudonat, C., Doroszewski, E., Gutschalk, A., 2025. Inter-Stimulus-Interval Dependence in Auditory Cortex Reflects Perceptual Organization of Streams Under Informational Masking. Psychophysiology 62, e70198. https://doi.org/10.1111/psyp.70198

Sadaghiani, S., Hesselmann, G., Kleinschmidt, A., 2009. Distributed and Antagonistic Contributions of Ongoing Activity Fluctuations to Auditory Stimulus Detection. J. Neurosci. 29, 13410–13417. https://doi.org/10.1523/JNEUROSCI.2592-09.2009

Schlossmacher, I., Dellert, T., Bruchmann, M., Straube, T., 2021. Dissociating neural correlates of consciousness and task relevance during auditory processing. NeuroImage 228, 117712. https://doi.org/10.1016/j.neuroimage.2020.117712

JNeurosci Accepted Manuscript

Schoenfeld, M.A., Hopf, J.-M., Merkel, C., Heinze, H.-J., Hillyard, S.A., 2014. Object-based attention involves the sequential activation of feature-specific cortical modules. Nat Neurosci 17, 619–624. https://doi.org/10.1038/nn.3656

Shafto, J.P., Pitts, M.A., 2015. Neural Signatures of Conscious Face Perception in an Inattentional Blindness Paradigm. Journal of Neuroscience 35, 10940–10948. https://doi.org/10.1523/JNEUROSCI.0145-15.2015

Shamma, S.A., Elhilali, M., Micheyl, C., 2011. Temporal coherence and attention in auditory scene analysis. Trends in Neurosciences 34, 114–123. https://doi.org/10.1016/j.tins.2010.11.002

Shinn-Cunningham, B.G., 2008. Object-based auditory and visual attention. Trends in Cognitive Sciences 12, 182–186. https://doi.org/10.1016/j.tics.2008.02.003

Supèr, H., Spekreijse, H., Lamme, V.A.F., 2001. Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). Nat Neurosci 4, 304–310. https://doi.org/10.1038/85170

Uusitalo, M.A., Ilmoniemi, R.J., 1997. Signal-space projection method for separating MEG or EEG into components. Med. Biol. Eng. Comput. 35, 135–140. https://doi.org/10.1007/BF02534144

Wiegand, K., Heiland, S., Uhlig, C.H., Dykstra, A.R., Gutschalk, A., 2018. Cortical networks for auditory detection with and without informational masking: Task effects and implications for conscious perception. NeuroImage 167, 178–190. https://doi.org/10.1016/j.neuroimage.2017.11.036

Wiens, S., Andersson, A., Gravenfors, J., 2023. Neural electrophysiological correlates of detection and identification awareness. Cogn Affect Behav Neurosci 23, 1303–1321. https://doi.org/10.3758/s13415-023-01120-5

Zhu, Y., Li, C., Hendry, C., Glass, J., Canseco-Gonzalez, E., Pitts, M.A., Dykstra, A.R., 2024. Isolating neural signatures of conscious speech perception with a no-report sine-wave speech paradigm. J. Neurosci. e0145232023. https://doi.org/10.1523/JNEUROSCI.0145-23.2023

Figure Captions

Figure 1. Schematic of the stimulus paradigm used. An isochronous repetition of six tones is presented as target amongst a multi-tone masker of random frequency tones, which are jittered relative to the onset of the target tones. In the stream-report task (SR), listeners indicate with

JNeurosci Accepted Manuscript

a button press as soon as they hear out the target tones as a distinct stream of isochronous tones. After each target masker combination, a cue tone is presented and listeners indicate in the frequency-report task (FR) if tones of this frequency were part of the previous stimulus set. (A) Example of a trial where the listener detects the target stream in SR and correctly confirms its frequency in FR (FR hit (SR+)). (B) Example of a trial where the target stream is not detected in SR but the frequency is correctly identified in FR (FR hit (SR-)). (C) Example of a catch trial with a correct rejection in both, SR and FR. (D) Example of a trial where the target stream is not detected in SR and also not identified after the cue tone in FR (FR miss (SR-)).

Figure 2. Behavioral results (mean ± standard error of the mean together with individual values, N=19). (A) Hit rates (Hit) plotted in darker colors and false-alarm rates (FA) in lighter colors. Stream report (SR, red) and all trials of the frequency report (FR, blue) are shown on the left . On the right side of the graph, FR results are shown for two subgroups stratified based on SR: SR- (yellow) includes FR Hit and FA rates where no response was made in SR. SR+ (green) includes FR Hit and FA rates where a response was already given in SR. (B) Detectability (d') and (C) Response criteria (c) for the same conditions and with the same color coding as in panel A.

Figure 3. Detected trials in the SR and FR tasks partitioned in 6 target-frequency bins (1: 501-603 Hz, 2: 661-794 Hz, 3: 871-1047 Hz, 4: 1148-1380, 5: 1514-1820 Hz, 6: 1995-2399 Hz). Colored bars indicate the mean across participants, error bars the standard-error of the mean, and black circles the single-participant data. (A) Absolute number of hit trials in the FR task for SR+ (left, green), SR- (middle, yellow), and miss trials (right, brown) by frequency bin. (B) Average trials number across frequencies for FR hit and miss trials. Full set based on behavioral data (behav.) on the left, trial numbers used for the MEG analysis (MEG) to the right. The latter excluded e.g. artefactual trials with too high amplitude in MEG. Datasets with less than five trials were excluded from the trial-wise MEG analysis (the two excluded are marked with an x). (C) Relative, trial-based hit rates for the SR task (percentage of all trials) and the FR task (relative to all SR- trials).

JNeurosci Accepted Manuscript

Figure 4. MEG data in auditory cortex (N=19) averaged relative to the tone-wise perception indicated in the streaming report (SR) task. (A) The regions of interest (ROI) in the auditory cortex used for the analysis are outlined in red. The ROI includes Heschl's gyrus and planum temporale in both hemispheres. (B) Grand-average source waveforms for the left and right auditory cortex for hits (red) and misses (thin, dark red) based on the ROI in (A). The interval shaded in darker gray represents the time interval 75–175 ms, in which the ARN amplitude was primarily measured. The subsequent interval shaded in lighter gray represents the time range 175–275 ms. (C) Average response amplitude in the time interval 75–175 ms (mean ± standard error of the mean; individual amplitudes overplotted). (D) Average response amplitude in the time interval 175–275 ms. Difference waveforms of this analysis are shown in Fig. S4A.

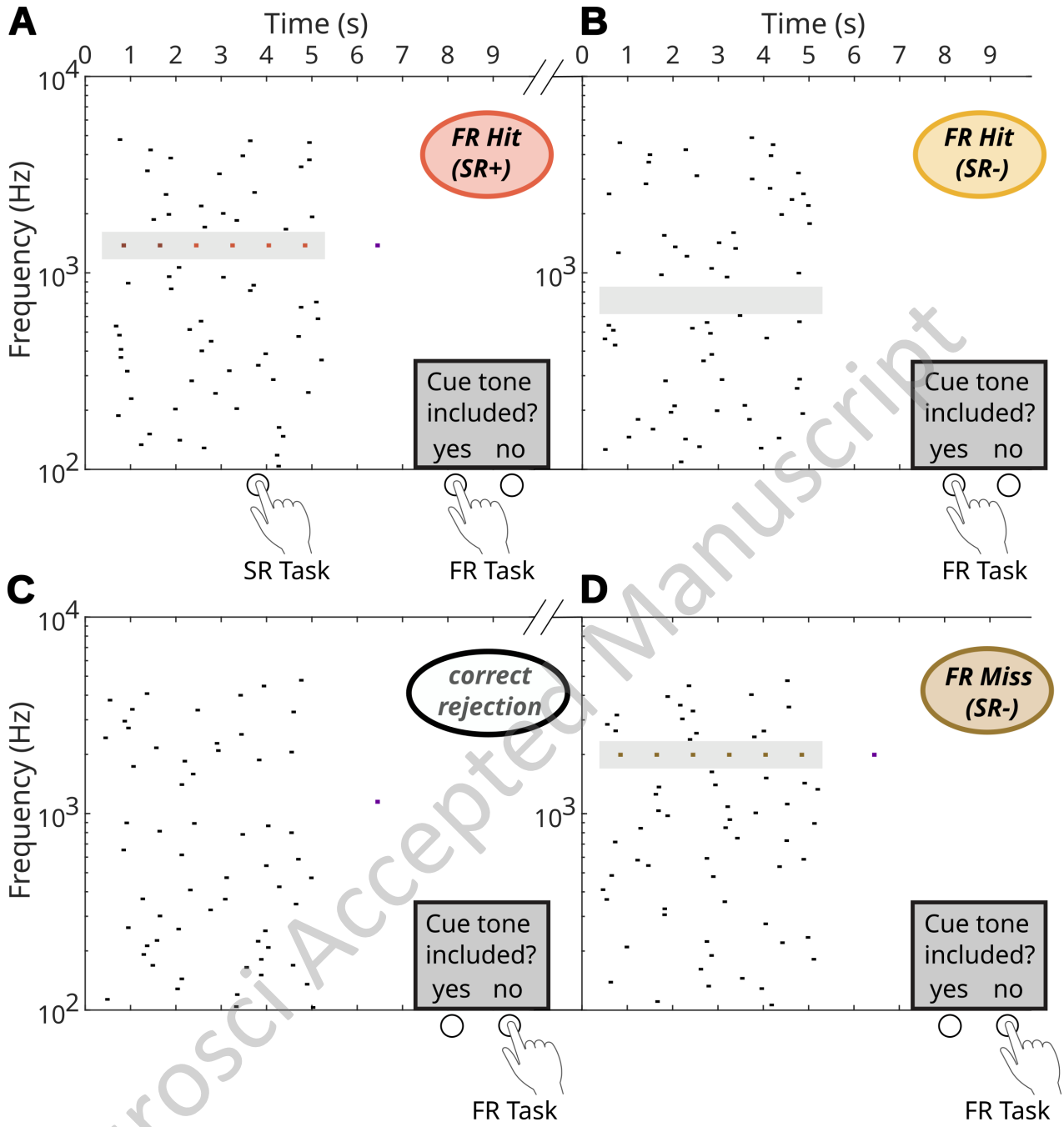
Figure 5. MEG data in auditory cortex averaged relative to the trial-wise perception in the SR and FR tasks (N=17). The data are based on the ROI shown in Fig. 4A. (A) Grand-average source waveforms for the left and right auditory cortex, for hits and misses in FR, sub-stratified for the response in SR. Green waveforms show hits in both, SR and FR (FR hits (SR+)), yellow show hits in FR not detected in SR (FR hits (SR-)), and thin brown lines indicate misses in SR and FR (FR misses (SR-)). The darker gray-shaded interval represents the time interval 75–175 ms, in which the ARN amplitude was primarily measured. The subsequent lighter gray interval represents the time range 175–275 ms. (B) Average response amplitude in the time interval 75–175 ms (mean ± standard error of the mean; individual amplitudes overplotted). Color scheme similar to panel A. (C) Average response amplitude in the time interval 175–275 ms. Difference waveforms of this analysis are shown in Fig. S4B.

Figure 6. Signal-detection models of neural activity based on the behavioral data (N=17). (A) Model 1 is based on the assumption that the different perceptions between FR hits that were detected (SR+) or missed in SR (SR-) represent a continuum that can be modeled by different response criteria within the Gauss distribution for signal trials shown in red; the noise distribution for no-signal trials is plotted in black. The response criterion between FR hits (SR+

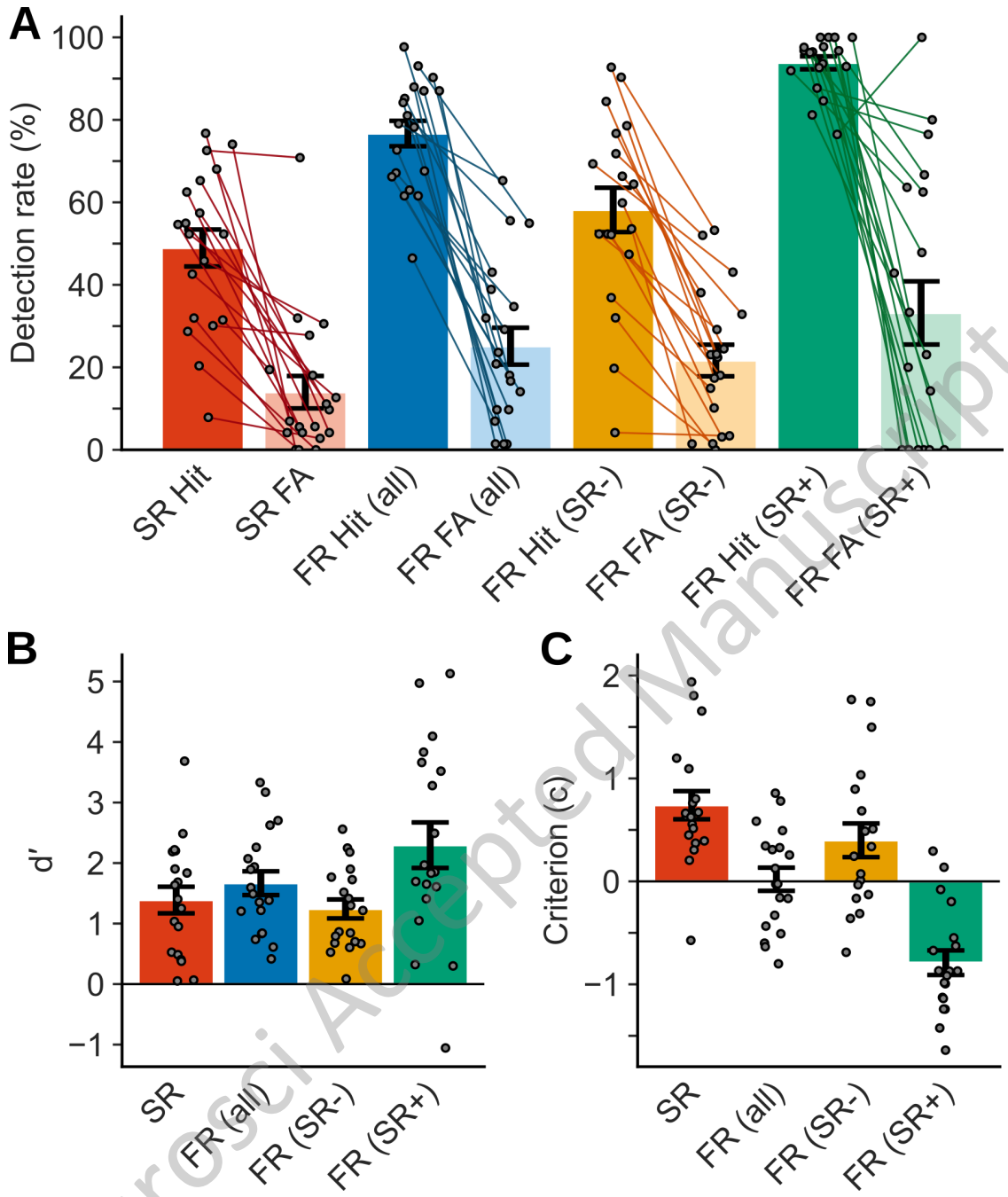
JNeurosci Accepted Manuscript

versus SR-) is indicated in green, between FR hits and FR misses (both SR-) in yellow. Note that the origin of signal strengths is set to the middle of the no-signal distribution, which is arbitrary. While the origin influences the absolute values of the estimates shown in panels B and C, it has no influence on the difference between these conditions. (B) Shows the average amplitude of the modeled neural activity, based on the integration over the intervals shown in A, for each individual data set (mean ± standard error of the mean together with individual values). Green: hits in SR and FR (FR hits (SR+)), yellow: FR hits missed in SR (FR hits (SR-)) and brown: misses in SR and FR (FR misses (SR-)). (C) Model 2 is instead based on the assumption that streaming and frequency perception are two distinct percepts, each represented in a separate Gaussian (yellow FR, green SR). The response criterion in the FR task is then modeled by a single criterion. (D) The internal signal strength predicted by this model for hits is estimated based on the signal to the right of the criterion for the respective gaussian. The signal for miss trials is based on the signal to the left of the criterion in the SR- gaussian (yellow).

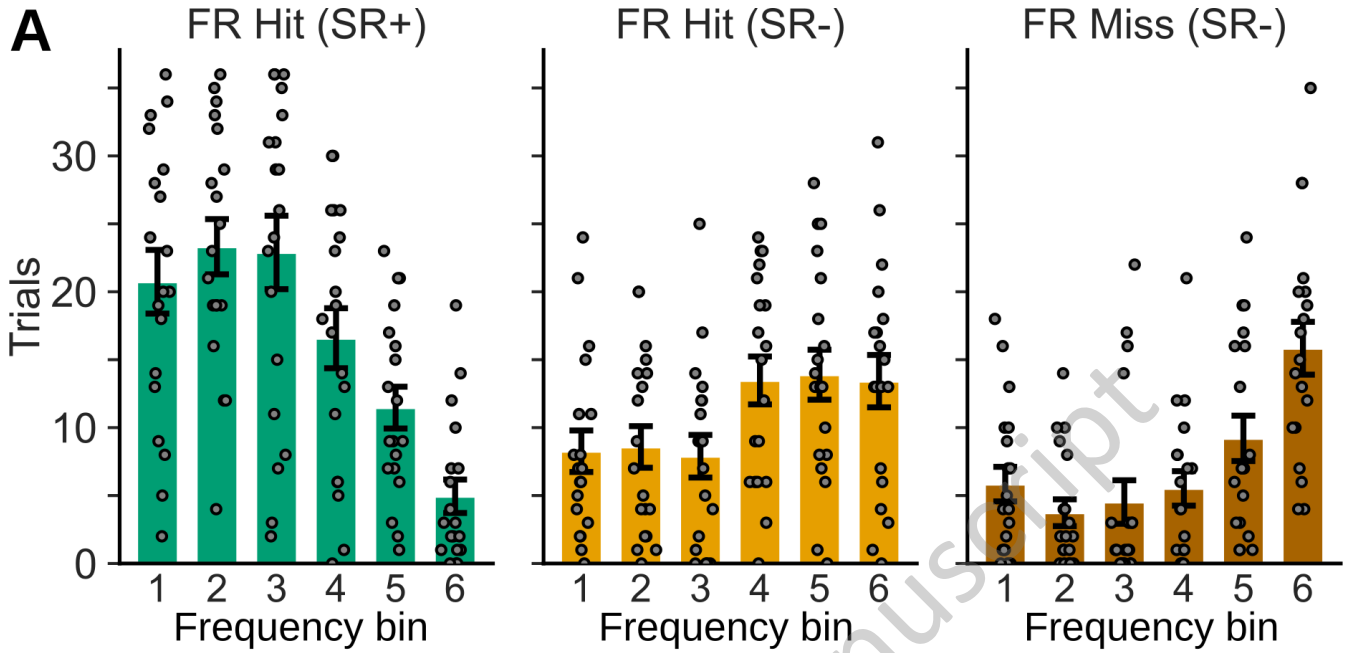
JNeurosci Accepted Manuscript

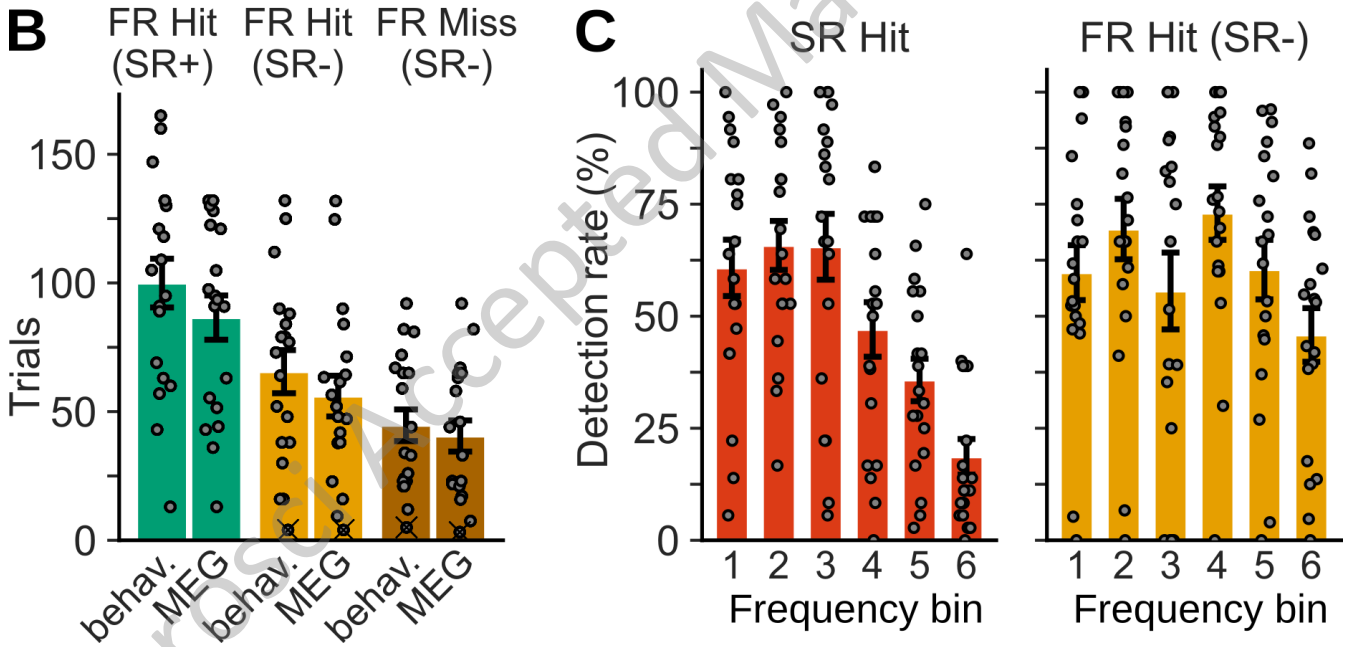


JNeurosci Accepted Manuscript

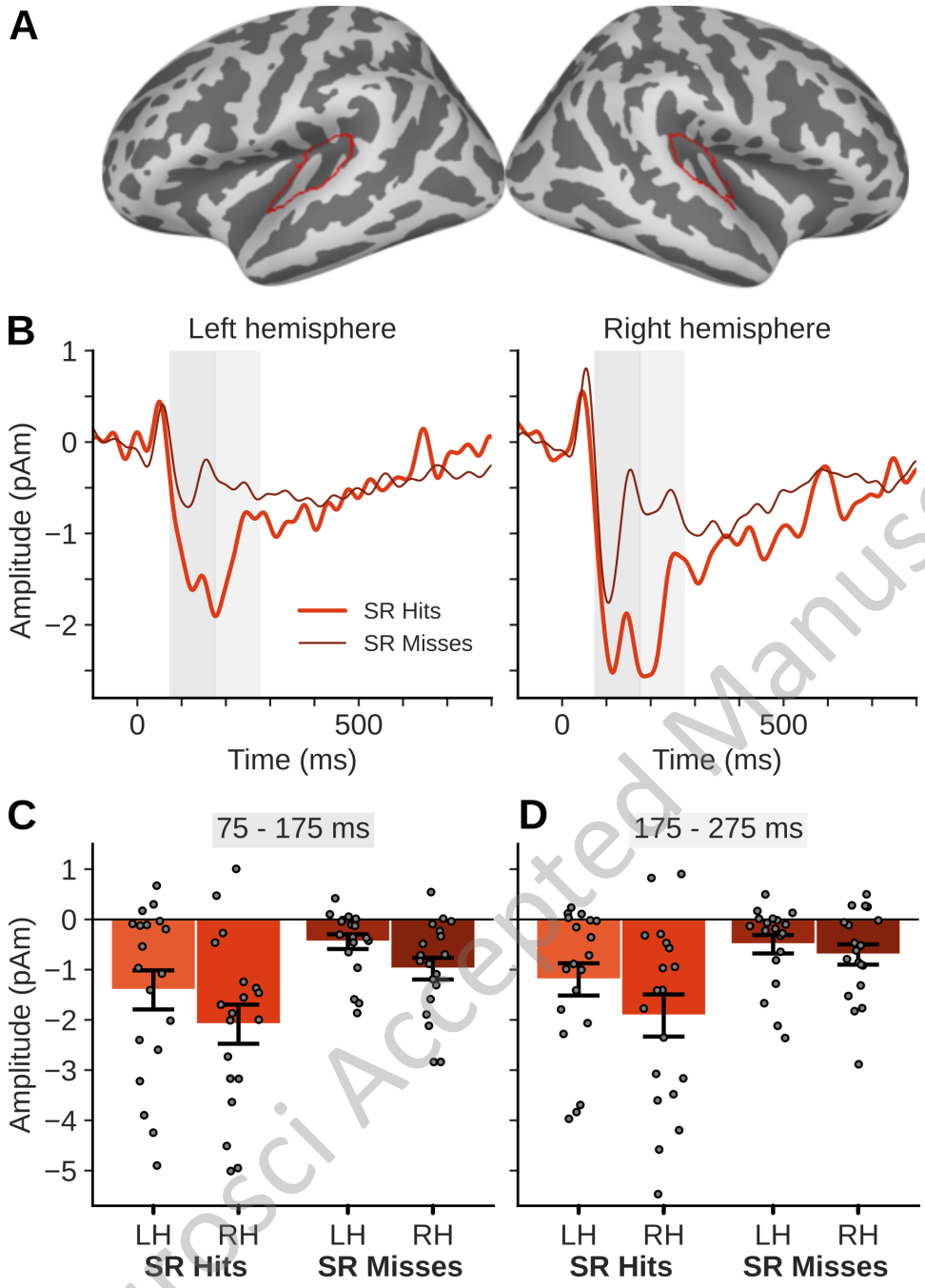


JNeurosci Accepted Manuscript

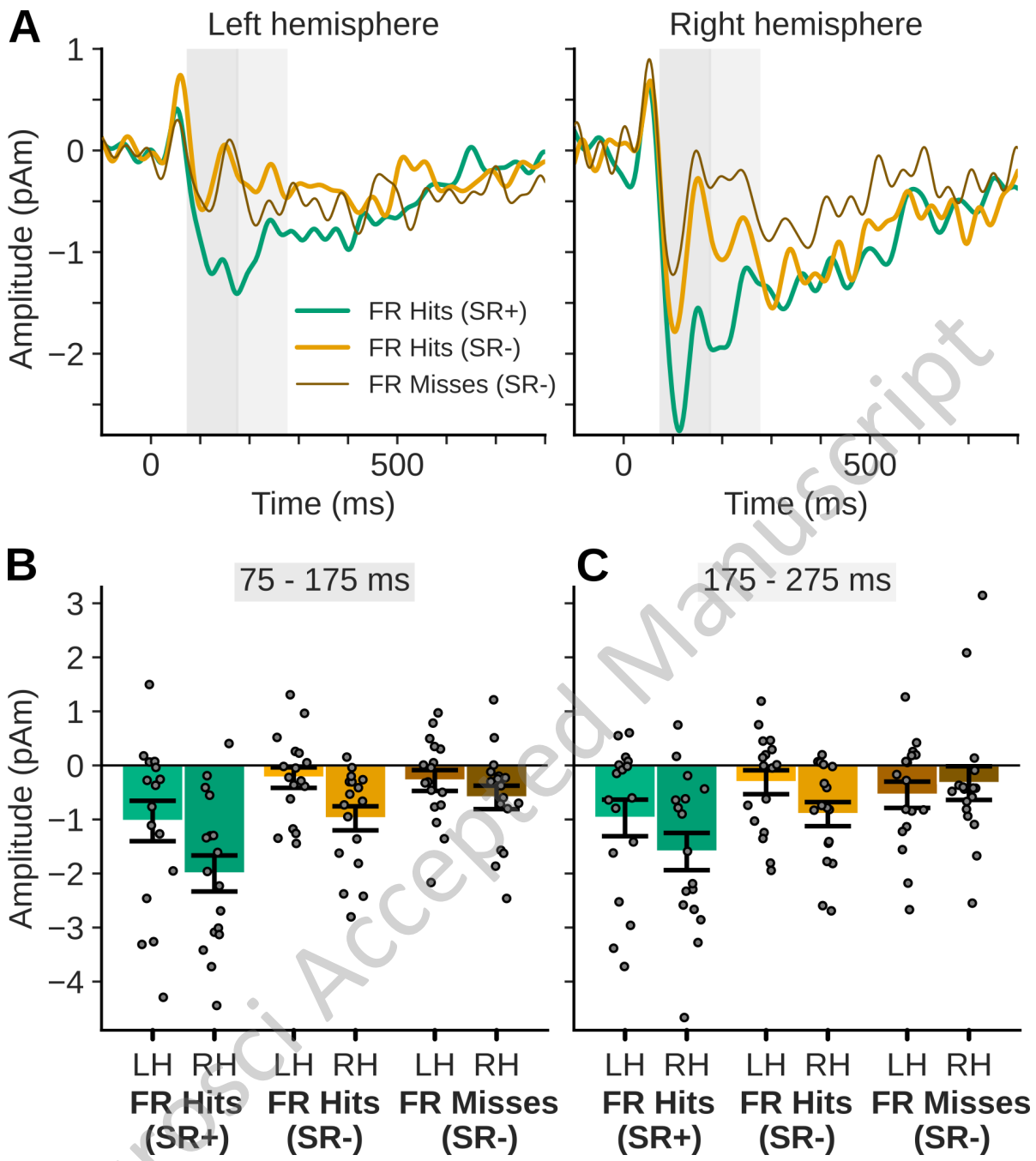




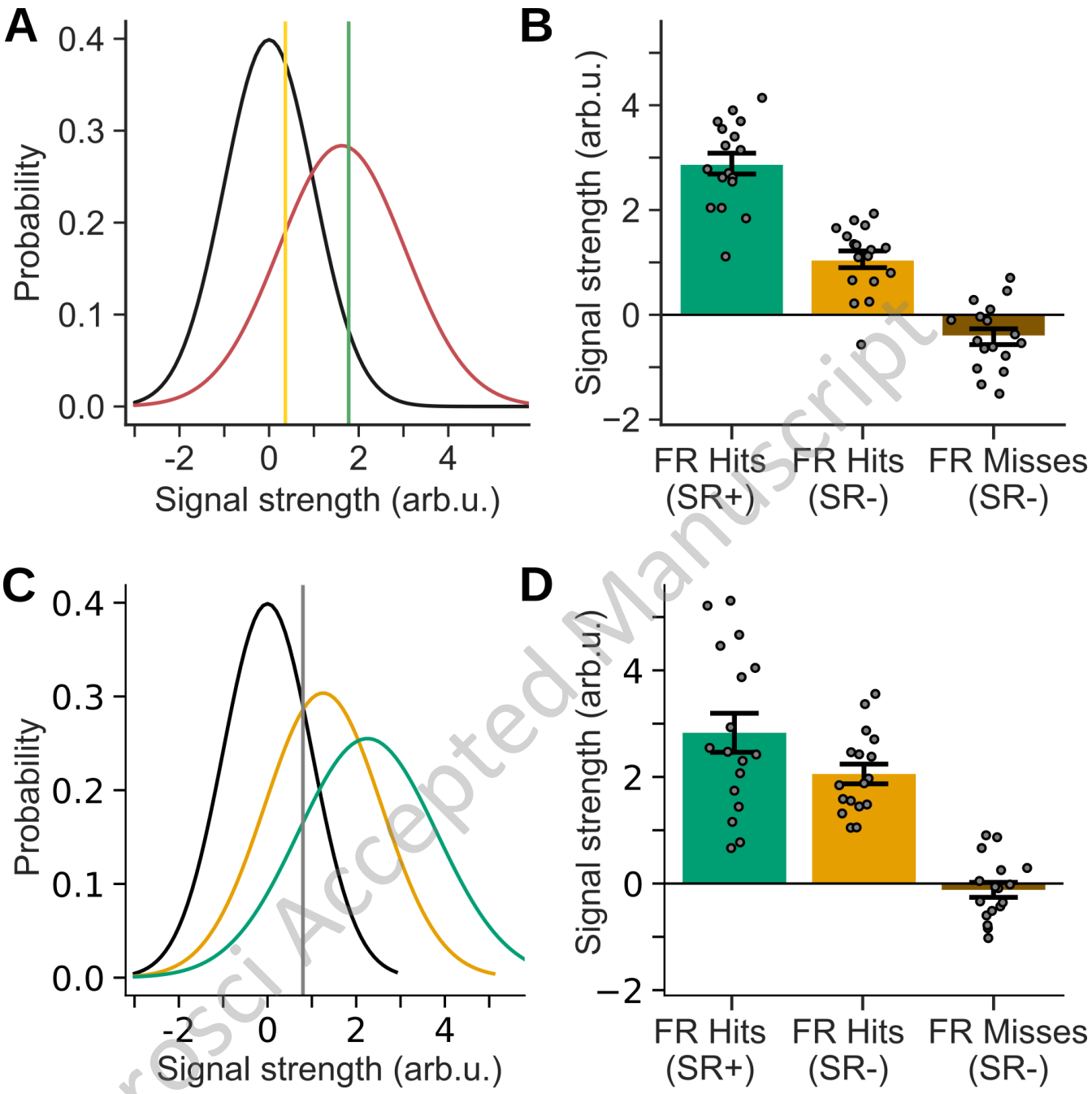
JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript