Auditory cortex activity related to perceptual awareness versus masking of tone sequences

Kai Gärtner, Alexander Gutschalk

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

ARTICLE Info

Keywords:
Conscious perception
Perceptual awareness
Auditory scene analysis
Magnetoencephalography
Informational masking
Attention

ABSTRACT

Sequences of repeating tones can be masked by other tones of different frequency. When these tone sequences are perceived, nevertheless, a prominent neural response in the auditory cortex is evoked by each tone of the sequence. When the targets are detected based on their isochrony, participants know that they are listening to the target once they detected it. To explore if the neural activity is more closely related to this detection task or to perceptual awareness, this magnetoencephalography (MEG) study used targets that could only be identified with cues provided after or before the masked target. In experiment 1, multiple mono-tone streams with jittered inter-stimulus interval were used, and the tone frequency of the target was indicated by a cue. Results showed no differential auditory cortex activity between hit and miss trials with post-stimulus cues. A late negative response for hit trials was only observed for pre-stimulus cues, suggesting a task-related component. Since experiment 1 provided no evidence for a link of a difference response with tone awareness, experiment 2 was planned to probe if detection of tone streams was linked to a difference response in auditory cortex. Random-tone sequences were presented in the presence of a multi-tone masker, and the sequence was repeated without masker thereafter. Results showed a prominent difference wave for hit compared to miss trials in experiment 2 evoked by targets in the presence of the masker. These results suggest that perceptual awareness of tone streams is linked to neural activity in auditory cortex.

1. Introduction

The presence of other sound sources can impair the perception of a target sound, even when the sounds do not overlap in their tonotopic representation in the cochlea (Kidd et al., 2008). This informational masking has e.g. been studied with single tones, presented together with multiple masker tones (Neff and Green, 1987). When the target is a sequence of repeated tones, masking is prominently reduced when the masker tones are newly randomized upon each target repetition, but not when the same masker is repeated together with the target (Kidd Jr. et al., 1994). This reduced masking has been ascribed to the perceptual grouping of the target tones into an auditory stream.

Multi-tone masking with repeated target tones is a powerful paradigm to study perceptual awareness, as the same physical stimulus can produce very different, salient percepts. Studies using magnetoencephalography (MEG) and electroencephalography (EEG) demonstrated that such repeated target tones evoke a negative-going wave in the auditory cortex when participants indicated that they are aware of the target stream, but not when it was masked (Giani et al., 2015; Gutschalk et al., 2008; Hausfeld et al., 2017). With reference to this experimental setup, this wave has been labeled the awareness related negativity (ARN) (Gutschalk et al., 2008). However, the exact relationship between the ARN and conscious auditory perception of the target tones has not yet been completely resolved (Dykstra et al., 2017; Snyder et al., 2015). In particular, it remains unclear if the ARN is related to the perceptual awareness of the single tones, or if it is more directly related to the awareness of the perceptual stream.

Another important aspect is the relationship between attention and the ARN. Attention is an umbrella term for multiple cognitive processes (Petersen and Posner, 2012). There is agreement that attention modulates the cortical processing of masked auditory scenes (Elhilali et al., 2009; Gutschalk et al., 2008; Molloy et al., 2019). Therefore, it is unlikely that the release of informational masking can be studied in the complete absence of attention. Some aspects of attention are more closely related to the task relevance of the target stimuli (Elgueda et al., 2019; Huang et al., 2019; Knyazeva et al., 2020), however, and one alternative interpretation of the ARN is that it reflects such task-related enhancement rather than perceptual awareness. The dissociation of these two aspects is difficult, in particular when listeners know the target and respond to it right away to indicate their perception of the ambiguous stimulus.

* Corresponding author.
E-mail address: Alexander.Gutschalk@med.uni-heidelberg.de (A. Gutschalk).

https://doi.org/10.1016/j.neuroimage.2020.117681
Received 23 September 2020; Received in revised form 24 November 2020; Accepted 15 December 2020
Available online 7 January 2021
1053-8119/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)
In the paradigms used in MEG so far, the target could be identified, first, by an identical tone repetition and, second, by the constant repetition interval that segregated it from the irregular intervals between target tones. In this setup, participants therefore immediately know that they detected the task-relevant tones, and it is possible that this task relevance then determines the processing of these tones in the auditory cortex.

Here, we probed the role of the immediate target identification on the generation of the ARN with a modified paradigm, where participants were informed about the target only after the target tones were presented under informational masking. To this end, we used two different stimulus configurations: in the first, we jittered the target repetition by the same amount as the masker tones, and repeated the masker tones instead of randomizing them newly in each repetition. Because the target could be identified solely based on tone frequency in this case, and thus based on a single tone, a second paradigm was developed, where the target-tone frequency was varied randomly, such that the target could only be identified when the masked target sequence was correctly matched to the unmasked sequence presented thereafter. Results showed that hit and miss trials only differed when participants correctly identified the whole target sequence, but not when they identified tone frequency based on post-stimulus cues.

2. Materials and methods

2.1. Participants

The study was approved by the ethical review board of Heidelberg University Medical School. All experiments were performed in accordance with the Declaration of Helsinki (2013 revision). Participants provided written informed consent and received payment for their participation. Exclusion criteria were any history of audiological, neurological, or psychiatric disorder, and magnetic implants that disturb MEG recordings.

In both experiments, a screening of the individual task performance was first obtained in a psychoacoustic test, typically on the day before the MEG recordings. Only listeners with $d' \geq 0.9$ were included for the MEG part of the respective experiment. In experiment 1, 37 listeners participated in the psychoacoustic test, of which 29 subsequently participated in the main MEG experiment. Nine of those were excluded from the analysis, because of insufficient signal-to-noise ratio of the MEG data ($n = 1$), insufficient task performance ($d' < 0.9$) ($n = 7$), or data loss ($n = 1$) within the MEG session. The remaining 20 participants (10 females; 4 left-handers; mean age: 24.5 years; range: 19–39y) were included in the analysis of experiment 1. In experiment 2, 20 listeners participated in the psychoacoustic test, and 15 of them were included in the main MEG experiment. Two participants were excluded, because of insufficient performance ($d' < 0.9$). Overall, 13 participants (4 females; one left-hander; mean age: 25y; range: 20–39y) were included in the analysis of experiment 2. Two of them had already participated in experiment 1.

2.2. Stimuli and experimental design

2.2.1. Experiment 1

The multi-tone clouds were based on 9 frequency bands, equally spaced on a logarithmic scale between 200 and 5000 Hz (frequency spacing corresponding to approximately 6.2 semitones). The target frequency was restricted to one of the three middle frequencies (699 Hz, 1000 Hz, or 1430 Hz) to avoid major frequency-dependent differences in detection rate observed for higher (and to some degree for lower) frequencies (Dykstra and Gutschalk, 2015; Gutschalk et al., 2008). Each of the remaining eight masker-tone frequencies were then randomized within their frequency bands in a range of ±2.5 semitones. The randomization was restricted such that a protected-frequency region of ±2/3 of an octave around the target tones was avoided by the masker tones. The masker tones next to the target were therefore generally chosen at the edge of the protected region. The multi-tone cloud comprised 5 tones of each target and masker frequency. Tone duration was 100 ms, including 5 ms long raised cosine ramps at the beginning and ending. The average inter-stimulus interval (ISI) was 500 ms (i.e. 600 ms SOA from onset to onset), and the tone onset of each target and masker tone was uniformly jittered by ±220 ms. There were 5 tones per target and 8 × 5 tones for the masker. In catch trials, 8 masker tones were configured around a protected region for one of the 3 target frequencies, but the target tones themselves were omitted. In the experiment, each target frequency was presented with equal probability with the constraint that more than two repetitions were not permitted. For the psychoacoustic screening test, 100 trials of the post-stimulus cue were presented; with 50% target and 50% catch trials. For the main MEG experiment, 216 trials were presented per set, including 25% catch trials. Each scene lasted 6120 ms with 1200 ms silence at the beginning and 120 ms silence at the end of each trial. The cue comprised two tones of the target frequency with a fixed 500 ms inter-stimulus interval and a minimum interval of 830 ms to the beginning/ending of the multi-tone cloud (Fig. 1A).

2.2.2. Experiment 2

The temporal configuration of the tone cloud was similar to experiment 1, but the frequency of both, target and masker tones, was changed for each 600-ms time interval. The frequency of the first target tone was randomly drawn from a logarithmic scale in the range from 400 to 2000 Hz. The following target tones were randomized in a range of ±2 semitones of the preceding tone, with the constraint that the randomization was repeated when the value fell out of the range 400–2000 Hz. The masker tones were randomly drawn from a logarithmic scale in the
range from 150 to 5000 Hz, excluding within each 600-ms time window a range of ±2/3 of an octave around the current target tone and ±2/3 of an octave from the directly preceding target tone. Six masker tones were drawn per time interval, with the additional constraint that each tone had a minimum distance of one semitone to the tones in the same and the previous time interval. The different randomization for target and masker tones was based on the assumption that perceived streams within a tone cloud typically comprise tones which are close by in frequency. If the same procedure would have been used for the masker tones, however, the overall seven streams would not have complied with the frequency differences between streams outlined above, at the same time. With the randomization used for the masker tones, the tone cloud was so dense that tones in subsequent intervals still had neighbors within a similar frequency range as in the tone stream, allowing for alternative lines to be heard by the listeners. The post-stimulus probe sequence replayed the target without the masker, after a minimum silent interval of 1160 ms duration. Two different kinds of catch trials were used, to control that listeners could not rely on the first or last tone of the target sequence, only. Standard catch trials did not comprise a target sequence, but instead a seventh masker tone randomized in the same way. The cue sequence was generated as described for the target sequences above. Alternative catch trials comprised a tone cloud with identical structure as described above. However, the cue sequence differed from the masked target in the three middle tones, while the first and last tones were retained. For the three middle tones, the tone steps were randomized in a 2-semitone range with the constraint that the frequency step was always in the opposite direction of the original sequence. The timing was identical to the masked target sequence. Overall, 100 trials were presented in the psychoacoustic test (70 target, 15 standard catch trials, and 15 alternative catch trials). For the main MEG experiment, two blocks of 200 trials were presented (70% target trials, 15% standard catch trials, and 15% alternative catch trials). The sampling rate was generally 48,000 Hz.

2.3. Procedures and data acquisition

Each listener started with a psychoacoustic test. At first, the task was explained and 20 practice trials were provided to get used to the task; this part could be repeated if required. Then the number of trials specified above were presented and the d’ was calculated. Only listeners whose detection and false alarm rate resulted in a sensitivity index $d' > 0.9$ were then asked to participate in the main MEG experiment. Stimuli were presented binaurally in a silent room with circumaural headphones (Sennheiser HDA200) with the sound intensity being adjusted to a comfortable listening level.

In the MEG experiment, sounds were generated and digital-to-analog transformed with an ADI-8 ds external audio interface (RME, Haimhausen) controlled with SoundMxPro software (HörTech, Oldenburg) in the Matlab environment. The analog signal was passed through programmable PA5 attenuators and then amplified with an HB7 headphone buffer (Tucker-Davis Technologies, Alachua) before being presented with ER3 headphones (Etymotic Research, Elk Grove Village) via 1-m-long plastic tubes and foam ear pieces. Stimuli were presented binaurally at a comfortable listening level.

After each trial, listeners were prompted via visual instruction to indicate if the cued target was present in the tone cloud or not, by pressing one of two buttons on an optical response box (Current Designs Inc.). The minimum inter-stimulus interval between two trials was 1300 ms in experiment 1 and 500 ms in experiment 2. The post stimulus pause was 500 ms, but in experiment 1 an additional 800 ms visual feedback was presented. In experiment 1, the first set used post-stimulus cues and the second set pre-stimulus cues, only. The average duration of each set was 32.4 min in both experiments.

The MEG was acquired with a Neuromag-122 whole-head system (MEGIN, Helsinki) with 122 planar gradiometers arranged in pairs at 61 positions. The data were recorded continuously with a sampling rate of 1000 Hz, direct coupled, and with a 330 Hz low-pass filter. Prior to recording, four head-position-indicator coils were fixed to the subjects’ head and digitized with reference to the nasion and two pre-auricular points, together with 100 additional points equally distributed over the head surface and face. The position of the coils relative to the dewar was then determined at the beginning of each recorded set.

2.4. Data processing and analysis

Data analysis was performed with BESA 5.1.8 (BESA GmbH, Gräfelfing, Germany). The target tones were averaged in an interval from 200 ms before to 400 ms after target-tone onset, subtracting a baseline in the interval 100 ms before to tone onset. Masked target tones and cue tones were separately averaged. When averaged in dependence of the behavior, all target tones in one trial were averaged together in the hit or miss category, respectively. Noise contaminated epochs were identified with the interactive “artifact scan tool”; rejection thresholds were individually adjusted, such that at least 90% of the intervals were accepted. Two dipoles (one in each hemisphere) were then fitted in a ~20 ms interval centered around the $N_1$ evoked by the cue tones in the pre cue condition (Fig. 2A). These dipoles were then applied to both sets as a spatial filter to obtain source waveforms for each condition. The spatial filter additionally included two regional sources in the position of the eyes to model artefacts caused by blinks and eye movements (Ille et al., 2002). Drifts and slow activity (caused e.g. by passing streetcars) were individually modeled for each condition by including the first component of a principal component analysis (PCA) in the time interval 200 ms before tone onset and 350–400 ms after tone onset. Source waveforms were low-pass filtered in MATLAB 2007 by applying a 2nd-order zero-phase shift Butterworth filter with a cut-off frequency of 20 Hz.

2.5. Statistical analysis

For statistical analysis of the MEG data, average amplitudes across the time interval 75–175 ms were measured in the source waveforms representing activity in left and right auditory cortex. Statistical analysis of target trials was based on an ANOVA for repeated measures with the factors detection (hit, miss) and hemisphere (right, left). The main hypothesis to be tested was that of a stronger negativity in the 75–175 ms time range for hit compared to miss trials, reflected by a main effect of detection in the ANOVA. The significance level was considered $p<0.05$ (two-tailed); the polarity of the change was separately confirmed in the plots presented. To test for priming effects on the post-stimulus cue tones, an ANOVA was calculated with the same parameters.

For the behavioral analysis, hit and miss rates were based on all trials (including those rejected for the MEG analysis). For the calculation of $d'$, we inserted $\text{Hit}=(n-0.5)/n$ in case of 100% hit rates, and $\text{Miss}=0.5/n$ was inserted in case of 0% false alarm rate, with n being the number of target-present or target-absent trials, respectively (Macmillan and Kaplan, 1985).

3. Results

3.1. Experiment 1

In experiment 1, both target and masker comprised repeated tones with the same temporal jitter (Fig. 1A). All participants were presented, first, with a set of stimuli where the tone frequency was indicated after the masked target (post cue) and, second, with a set where the tone frequency was indicated before the masked sequence (pre cue). Performance was prominently better for pre- compared to post-stimulus cues (Fig. 1C) with higher hit rates (97.04% vs 75.99%; $F_{1,19} = 54.02; P < 0.001$) and lower false-alarm rates (4.17% vs 16.30%; $F_{1,19} = 33.88; P < 0.001$). The comparison of target frequencies reveals that the middle frequency was somewhat more frequently detected in the post-cue condition (mean hit rates: 699Hz: 69.72%; 1000Hz: 84.54%; 1430Hz:
Fig. 2. Neural activity in auditory cortex evoked by target and cue tones. (A) Mean dipole positions in approximated Talairach coordinates plotted on a standard brain \( (N = 20); \text{LAC: } x = -48 \pm 5, y = -24 \pm 6, z = 5 \pm 10; \text{RAC: } x = 47 \pm 6, y = -20 \pm 10, z = 6 \pm 8; \text{mean } \pm \text{ standard deviation}. \) (B) Grand-average source waveforms evoked by the target tones in the post-stimulus (upper part) and the pre-stimulus cue conditions (lower part). Waveforms for hit (green) and miss (orange) trials are plotted on top of each other. \( \text{(C) } \) Mean amplitude (± standard error, \( N = 20 \)) in the time interval 75 ms–175 ms. The data of the post-cue condition is plotted with closed, the data of the pre-cue condition with open symbols. \( \text{(D) } \) Difference (black) between pre-cue hit (dark green) and post-cue hit trials (light green) for left (solid) and right AC (dashed). \( \text{(E) } \) Grand-average source waveforms evoked by post-cue tones for detected (green) and undetected (orange) trials. \( \text{(F) } \) Mean amplitude of the response evoked by post-cue tones (± standard error, \( N = 20 \)) in the time interval 75 ms–175 ms.

Table 1.

| Test                                | \( F \) (1,19) | \( p \)  |
|-------------------------------------|----------------|--------|
| **Post-stimulus cue**               |                |        |
| detection                           | 0.48           | 0.4948 |
| hemisphere                          | 0.69           | 0.4177 |
| detection’hemisphere                | 0.06           | 0.8037 |
| **Pre-stimulus cue**                |                |        |
| detection                           | 9.29           | 0.0066 |
| hemisphere                          | 0.69           | 0.4177 |
| detection’hemisphere                | 0.19           | 0.6650 |
| **Hit trials (pre- versus post-stimulus cue)** | | |
| cue-position                        | 14.04          | 0.0014 |
| hemisphere                          | 0.18           | 0.6378 |
| cue-position’hemisphere             | 0.69           | 0.4181 |

73.70%; main effect of frequency: \( F_{2,18} = 10.20; P = 0.0011 \), but not in the pre-cue condition (mean hit rates: 699Hz: 98.15%; 1000Hz: 97.87%; 1430Hz: 95.09%; main effect of frequency: \( F_{2,18} = 3.38; P = 0.0567 \)).

Dipole source waveforms for the target tones \( \text{(Fig. 2B, C)} \) were estimated with dipoles in the left and right auditory cortex, fitted to the \( N_1 \) evoked by the unmasked target tones \( \text{(Fig. 2A)} \). Statistical results are summarized in Table 1: no difference in source activity was observed between hit and miss trials in the post-cue condition. For the pre-cue condition, in contrast, negative-going source activity was longer and significantly higher for hit compared to miss trials. Note that the source activity for miss trials is strongly overlaid with alpha activity because of the small number of trials. Source amplitudes for hit trials were also significantly larger in the pre-cue condition compared to the post-cue condition. A closer comparison between the overlaid waveforms \( \text{(Fig. 2D)} \) suggests that this difference emerges around and after the peak of the \( N_1 \) m, which is clearly recognized only in the post-cue condition, whereas the onset up to the \( N_1 \) m peak is very similar. It therefore appears that the difference between pre- and post-cue hit trials is mostly caused by a separate long-latency component subsequent to the \( N_1 \) m, with a peak latency around 180 ms.

We also compared the response evoked by the post-stimulus cue tones between hit and miss trials \( \text{(Fig. 2E and F)} \), to explore if a priming effect was observed for detected targets. Results showed a stronger negative-going response for hit compared to miss trials (deection: \( F_{1,19} = 5.61; P = 0.0286 \); detection × hemisphere: \( F_{1,19} = 1.28; P = 0.2714 \)).

The results of the post-cue condition in experiment 1 argue against an interpretation of the ARN as a neural correlate of the awareness of single tones under informational masking. One possibility, suggested by the pre-cue condition, is that enhanced negative-going activity is merely an indication of task-related attention. Alternatively, it could be that the pre-cue promotes the perception of the target frequency into one
coherent stream, concurrently enhancing behavioral performance in the detection task. The latter interpretation cannot be tested with the data from experiment 1; it would predict that an ARN can be observed with post-stimulus cues, provided that the task requires that major parts of the target stream are perceived, rather than just single tones of that stream, as was sufficient for experiment 1.

3.2. Experiment 2

To probe the hypothesis that the ARN is related to the awareness of auditory streams in experiment 2, random tone sequences were used as targets, which were newly randomized upon each trial and therefore unknown to participants when presented within the multi-tone masker (Fig. 3). Only post-stimulus cues were used in experiment 2, which replayed the complete target sequence. Whereas the hit rates were lower in comparison to both conditions of the first experiment (mean hit rate: 47.31%), the false-alarm rates were in the same range as in the pre-cue condition of experiment 1 (3.46%) when no target was present, leading to an average d’ of 1.783. In a subset of trials, a target was present but a different sequence was used as post-stimulus cue. This post-stimulus cue adopted the first and last tone from the target sequence, but comprised three different tones in between. These trials were included to control for the possibility that the sequence could be identified by a subset of tones, in particular by the first or last tones; the false alarm rate was 9.10% in this case, and the difference between false-alarm rates for target-absent and different-sequence trials was highly significant (F(1,13) = 26.61; p = 0.0002). This finding suggests that the presence of a target-like stream makes the correct rejection more difficult than when no target-like stream is present, but the false alarm rate for target-like streams was still much lower than the hit rate for correct targets.

MEG source waveforms evoked by the tones of the target streams are shown in Fig. 4B. Evoked responses are prominently enhanced for hit compared to miss trials (Fig. 4C, Table 2). The response to correct rejections of different-sequence trials lies in between hit and miss trials. This is consistent with the assumption that the tone sequence embedded in the masker was perceived in about half of the trials, but that the correct behavioral response (rejection) does not dissociate between trials where the tone sequence was heard and those where it was not heard.

In contrast to experiment 1, the comparison of responses evoked by post-stimulus cue tones revealed no significant effect of detec-
Table 2: Statistical results of experiment 2 for target-evoked amplitudes in the source waveforms.

| Test                        | F (1,12) | p     |
|-----------------------------|----------|-------|
| Detection                   | 20.5     | 0.0007|
| Hemisphere                  | 2.61     | 0.1319|
| Detection × Hemisphere      | 2.37     | 0.1499|
| Detection by tone position (75–175 ms) | |     |
| 1. Tone:                    | 3.39     | 0.0906|
| 2. Tone:                    | 31.39    | 0.0001|
| 3. Tone:                    | 20.37    | 0.0007|
| 4. Tone:                    | 7.82     | 0.0162|
| 5. Tone:                    | 7.32     | 0.0191|
| Detection by tone position (175–275 ms) | |     |
| 1. Tone:                    | 0.41     | 0.5333|
| 2. Tone:                    | 14.97    | 0.0022|
| 3. Tone:                    | 7.37     | 0.0188|
| 4. Tone:                    | 9.24     | 0.0103|
| 5. Tone:                    | 16.21    | 0.0017|

4. Discussion

The results of experiment 2 of this study show that the long-latency negativity in auditory cortex, referred to here as the ARN (Gutschalk et al., 2008), covaries with perceptual awareness of random-tone streams in a multi-tone masker background. In contrast, experiment 1 did not provide such evidence of a difference response between hit and miss trials in a setting where the target was identified based on its frequency provided by a post-stimulus cue. Since this task did not require that the target tones are perceived as segregated stream, experiment 1 suggests that perceptual awareness of a tone’s presence inside of a multi-tone masker is not necessarily coupled to ARN. The difference response observed for the pre-cue condition fulfills criteria for task-related attentional enhancement; it may also be associated with modulation of the perceptual organization, but this cannot be determined in retrospect.

In contrast to previous studies, participants did not know if they had detected a target while listening to the masked stimulus before the post-stimulus cue. In experiment 1, there were multiple mono-tone streams with random onset. In experiment 2, the target-tone sequence was random but with a maximum distance of two semitones between directly subsequent tones, and a minimum distance of four semitones between target and masker tones in the same time interval. The frequency distance between target and masker tones, which is one important parameter for segregating a tone sequence from a multi-tone masker (Micheyl et al., 2007), was similar in both experiments. However, the similarity between target and masker structure (Durlach et al., 2003) was lower in experiment 2, because the randomization was less constrained for masker tones. It is therefore possible that the target was segregated more easily in experiment 2, however, no behavioral data regarding the perception of auditory streams is available for comparison from experiment 1.

4.1. Tone versus tone-pattern masking

While target tones were usually not cued in traditional informational masking paradigms, the same, known frequency was typically used (Kidd Jr. et al., 1994; Neff and Green, 1987), which allows for full focus of attention to that frequency. At least for the stimuli in experiment 1 of the present study, the question may therefore be raised, if missed target tones were really “masked”, i.e. not perceived, or if the target tones are simply difficult to single out from the cloud in retrospect, when their frequency is not known beforehand. Miss trials in the post-cue condition of experiment 1 might then be caused by a limitation of short-term memory rather than a lack of perception. Conversely, the higher performance in the pre-cue condition could be readily explained by the lower memory load required for this task. The enhanced negativity in the pre-cue task additionally demonstrates that the target tones are differently processed after pre-cues. Whether this also goes along with a different perception or perceptual organization cannot be dissociated based on the reporting task used. In contrast, we consider it very unlikely that the target-tone patterns in experiment 2 could be recognized in the context of the masker tones without being organized into a distinct perceptual stream during perception and encoding in short-term memory, already. The computation load would here be much higher than for the recognition of a single tone, which was already difficult.

Based on these considerations, we suggest that there are two levels of informational masking: First, there is the masking that has been traditionally described with multi-tone maskers, where single (or repeated) tones are not perceived at all (Kidd Jr. et al., 1994; Neff and Green, 1987). Second, the higher-level quality of a stream, which is readily perceived without masker, can separately be masked, without necessarily masking the single tones comprised in the stream. We suggest that such tone-pattern masking dominates in experiment 2 of this study. Based on these data, one possible interpretation of the ARN is that it reflects the perception of streams rather than of single tones.
Tone-pattern masking appears as the opposite site of stream (or figure) formation from a random-tone background (Michel et al., 2007; Teki et al., 2011), and it has been pointed out that a major part of e.g. speech masking is likely related to the disruption of sequential stream formation (Kidd et al., 2008). While a listener typically remains aware of all tones in a classical two-tone streaming paradigm (Van Noorden, 1975), hearing out a stream from a multi-tone masker typically coincides with the perceptual awareness of the tones (Michel et al., 2007). If or how these tones are perceived when the target stream is masked remains difficult to explore. In particular when the masker tones are presented synchronously (Kidd et al., 1994; Michel et al., 2007; Teki et al., 2011), the masking is not only explained by the lack of stream formation, but also by the alternative grouping of the tones into serial chords. The temporal jitter used in this and previous studies (Elhilali et al., 2009; Gutschalk et al., 2008) reduces such grouping and allows for easier glimpsing on the single tones (Demany et al., 2011), which allows for hearing out single tones more easily in the absence of stream formation, as we think is the case for target tones in experiment 1 of this study.

Still, the informational masking of tones and streams may often co-occur. In the present study, tones evoked a clear N1 when the tone pattern was not detected, and a larger, more broad-based ARN in detected trials. Other studies with similar multi-tone maskers and softer target tones did not observe N1 for undetected trials (Dykstra and Gutschalk, 2015; Gutschalk et al., 2008), and similarly N1 was only observed for detected but not for missed tones in noise (Hillyard et al., 1971). Possibly, ARN related processes code for tone context, which can reach from a solitary event to one tone amongst multiple similar in a multi-tone cloud, reflecting a continuum from salient to masked. Thus, when a tone (or speech) pattern is masked, its rhythm, melody, (or prosody) may disappear, but the tone may still be perceived in a multi-tone context, and evoke a small N1 as in the present study.

4.2. Interaction of attention and perceptual organisation

The difference between hit trials in the pre-cue and post-cue conditions of experiment 1 could simply be explained by attention along the tonotopic axis (Riecke et al., 2018) that evokes an additional component after the N1, previously referred to as negative difference wave N2 (Hansen and Hillyard, 1980; Rif et al., 1991) or processing negativity (Nätänen et al., 1978). However, it has been suggested that the N2 already operates on auditory streams rather than on pure tonotopy (Alain and Woods, 1994). This interpretation has received further support by similar, negative-going response enhancement observed for attended speech streams (Ding et al., 2014; Ding and Simon, 2012; Power et al., 2012) and active versus passive listening to stochastic figure-ground stimuli (O’Sullivan et al., 2015). Along these lines, the longer-latency negativity for pre-cue hits could mean that the target tones were organized into a perceptual stream. More generally, it is possible that this N2 is the longer-latency component of the ARN, which is related to active listening to auditory streams.

In experiment 2 and in earlier studies, the ARN additionally included the latency range of the N1 (Dykstra and Gutschalk, 2015; Gutschalk et al., 2008), and the earlier part of the ARN showed N1-like behavior with respect to stimulus lateralization (Königs and Gutschalk, 2012). We therefore expect that the N1 part of the ARN is also related to the if and how a stream of tones is perceived. The build up of the response enhancement for hit trials in experiment 2 shows that the longer-latency part of the ARN is only present from the second tone on, but the difference between hit and miss trials is driven by both time intervals, as the N1 decreases for miss trials from the second tone on. Previous studies observed a different build-up pattern, where the second (and subsequent), but not the first tone of the sequence evoked an early ARN (Giani et al., 2015; Wiegand and Gutschalk, 2012). The difference is probably related to the target onset coinciding with the masker onset in the present experiments, whereas the masker started before the targets in the previous studies, and the response was therefore adapted, already. Once the target is perceived as separate stream, however, ARN appears to partly re-adapt to the longer time intervals of the target stream, instead, similar as suggested for streaming of two-tone sequences (Gutschalk et al., 2005).

Overall, the representation of auditory streams in different background conditions remains closely coupled to selective attention, and two different concepts have been entertained for the negative difference response: one is response enhancement by focus of attention (Ding and Simon, 2012; Elhilali et al., 2009; Hansen and Hillyard, 1980), the other, ARN, assumes a direct relationship to perceptual awareness (Gutschalk et al., 2008; Snyder et al., 2015). The question then remains, which of the two concepts captures the difference response better: if the response was required for perceptual awareness of a stream, the stream should only be perceived when this type of response is evoked, but not otherwise. If the response reflects enhancement by attention, then the stream may already be perceived before, and the perception may or may not change as the response is enhanced by attention. This dissociation has not been successfully tackled today, and a third variant was in fact that the two cannot be dissociated at all (Posner, 1994). One argument against a purely attention-based, task-related interpretation is that similar activity in secondary auditory cortex is observed when a single stream is presented in a passive setting (Dykstra et al., 2016; Gutschalk et al., 2008), when no major activity is observed in areas outside of the auditory cortex (Wiegand et al., 2018). While it has been shown that distraction of attention to another modality can further reduce the N1 in such a setting, the response was still present even for high visual loads (Molloy et al., 2015). We therefore suggest that this coupling of attention and perceptual awareness is in particular found in situations of perceptual competition, where multiple perceptual interpretations exist (Desimone and Duncan, 1995). However, we suggest that the same network in auditory cortex may be active during the perception of a single stream in silence, without requirement for selective attention, and therefore suggest that the ARN requires attention in the experiments presented here, but is at the same time closely coupled to the perceptual awareness of auditory streams.

Data and code availability statement

Single-subject source level data and matlab readers are available on heiDATA (https://doi.org/10.11588/data/Y8UEOY).

Declaration of Competing Interest

The authors declare no competing financial interests.

Credit authorship contribution statement

Kai Gärtner: Conceptualization, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. Alexander Gutschalk: Conceptualization, Formal analysis, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

This work was supported by Deutsche Forschungsgemeinschaft [Grant DFG 593/5-1].

References

Alain, C., Woods, D.L., 1994. Signal clustering modulates auditory cortical activity in humans. Percept Psychophys 56, 501–516.

Demany, L., Semal, C., Pressnitzer, D., 2011. Implicit versus explicit frequency comparison: two mechanisms of auditory change detection. J. Exp. Psychol. Hum. Percept. Perform. 37, 597–605. doi:10.1037/a0020368.

Desimone, R., Duncan, J., 1995. Neural Mechanisms of Selective Visual. Annu. Rev. Neurosci. 18, 193–222.
