Membrane potential dynamics of populations of cortical neurons during auditory streaming

Brandon J. Farley\(^2\) and Arnaud J. Noreña\(^1\)

\(^1\)Aix Marseille Université, Centre National de la Recherche Scientifique, Marseille, France; and \(^2\)Neurology Research, Biogen, Cambridge, Massachusetts

Submitted 3 June 2015; accepted in final form 12 August 2015

Farley BJ, Noreña AJ. Membrane potential dynamics of populations of cortical neurons during auditory streaming. J Neurophysiol 114: 2418–2430, 2015. First published August 12, 2015; doi:10.1152/jn.00545.2015.—How a mixture of acoustic sources is perceptually organized into discrete auditory objects remains unclear. One current hypothesis postulates that perceptual segregation of different sources is related to the spatiotemporal separation of cortical responses induced by each acoustic source or stream. In the present study, the dynamics of subthreshold membrane potential activity were measured across the entire tonotopic axis of the rodent primary auditory cortex during the auditory streaming paradigm using voltage-sensitive dye imaging. Consistent with the proposed hypothesis, we observed enhanced spatiotemporal segregation of cortical responses to alternating tone sequences as their frequency separation or presentation rate was increased, both manipulations known to promote stream segregation. However, across most streaming paradigm conditions tested, a substantial cortical region maintaining a response to both tones coexisted with more peripheral cortical regions responding more selectively to one of them. We propose that these coexisting subthreshold representation types could provide neural substrates to support the flexible switching between the integrated and segregated streaming percepts.

auditory scene analysis; cocktail party; perceptual segregation; bistability; neural adaptation

A FUNDAMENTAL ASPECT OF HEARING is being able to segregate and attend to specific auditory objects despite the mixed acoustic input reaching the ears, as exemplified in the “cocktail party problem” (Cherry 1953). The solution requires appropriately integrating acoustic input that comes from a single source and segregating input originating from different sources.

The canonical auditory streaming paradigm is a model used to study the problem of “sequential segregation,” that is, integrating or segregating the important class of auditory objects that have a pseudorhythmic structure, such as speech, other communication signals, or music (Berrymen 1976; Bregman 1990; Ghitza and Greenberg 2009). In this paradigm, two pure tones are played alternately and are perceptually heard as one integrated or two segregated “streams,” based on key parameters such as the frequency difference between them and their presentation rate (van Noorden 1977). The dependence of segregation on these parameters presumably derives from the brain’s assumptions about what constitutes probable sound sources (Bregman 1990), similarly to Gestalt principles in vision (Rock and Palmer 1990). How and where in the auditory system these principles are implemented mechanistically remain unclear.

Neurophysiological studies of suprathreshold (spiking) activity in single neurons have been performed during the auditory streaming paradigm. Based on recordings at individual sites, population activity patterns have been inferred and used to motivate the hypothesis that perceptual segregation of different sources is related to the spatial separation of cortical representations induced by each acoustic source or stream (Fishman et al. 2001; Micheyl et al. 2005). An alternative hypothesis is that two streams form when two temporally incoherent neural assemblies arise (Elhilali et al. 2009). The second hypothesis can be considered an extension of the first by considering both the temporal and spatial representations of acoustic stimuli.

Each of these hypotheses, relating to patterns of spiking activity, can account for various properties of stream segregation such as its dependence on the frequency separation between the tone sequences and their presentation rate, as well as the “buildup” of stream segregation over time. However, the streaming percept is subject to modulatory influences, in addition to bottom-up acoustic input. For example, it has been shown that the percept can alternate over time between one and two streams (i.e., “ bistability”) despite an unchanging pattern of acoustic input (Denham and Winkler 2006; Pressnitzer and Heupe 2006; Schwartz et al. 2012). The neural mechanisms that permit this flexible switching between perceptual alternatives remain unclear and may involve processes other than spiking patterns.

In addition to spiking activity, the pattern of subthreshold excitability has been proposed to contribute to the perceptual organization of sensory responses (Lakatos et al. 2008). Indeed, given the broad range of input represented in subthreshold compared with suprathreshold activity (e.g., the spectral integration is much broader for subthreshold activity), it seems plausible that the two signals may play different roles in the context of auditory scene analysis (Carandini and Ferster 2000; Chadderton et al. 2009; Galvan et al. 2002; Moore et al. 1999; Noreña and Eggermont 2002). Local field potentials, which are sensitive to subthreshold excitability, have previously been shown to provide clues about the mechanisms of auditory selective attention (Lakatos et al. 2013). However, the role that subthreshold activity plays in sound integration vs. segregation, as measured in the auditory streaming paradigm, has not yet been examined.

To address this issue we used, for the first time, voltage-sensitive dye imaging (VSDI) (Shoham et al. 1999) to measure cortical neural activity patterns during the auditory streaming
paradigm. Previous studies combining VSDI with whole-cell voltage recordings have established that when applied to the mammalian cortex in vivo, VSDI measures a population subthreshold membrane potential-related signal (as opposed to spiking activity) from neural elements within the superficial layers of the cortex (Peterson et al. 2003). VSDI thus permits the measuring of subthreshold membrane potential-related activity at high spatial and temporal resolution across the auditory cortex (Farley and Noreña 2013).

Our results, based on using this technique, are in broad agreement with previous studies showing that the spatiotemporal separation of cortical responses to alternating tone sequences parallels the tendency to segregate perceptually the acoustic sequences. In addition, however, we find that across most streaming paradigm conditions tested, a substantial cortical region maintaining a subthreshold membrane potential response to both sequences coexists with more peripheral cortical regions responding more selectively to one of them. We propose that the spatiotemporal pattern of the cortical subthreshold membrane potential may provide a substrate for the flexibility (e.g., bistability) of the auditory streaming percept.

METHODS

Surgical procedures. Experimental protocols followed the guidelines of French and European laws on laboratory animal care and use and were approved by the Animal Care Committee of Bouches-du-Rhône, France. The data from this study come from 12 male and female adult guinea pigs (Dunkin Hartley breed from Charles River Laboratories, Wilmington, MA). Experimental procedures for VSDI have been described previously (Farley and Noreña 2013). Immediately before surgery, each animal was checked for the presence of a bilateral pinnal reflex in response to sound. Animals were pretreated with atropine, and anesthesia was induced and maintained with a mixture of ketamine (46 mg/kg) and xylazine (24 mg/kg), with supplemental doses at one-half of this concentration each hour. Lidocaine was applied to incision points. Body temperature was maintained at 37°C. A tracheotomy was performed, and animals were artificially respired with air containing concentrated oxygen. Anesthesia depth was assessed by monitoring the ECG rate and expired CO₂ levels. Animals were immobilized by a metal piece cemented to the skull, which could also be fixed to a stereotaxic apparatus. A craniotomy and durotomy were performed unilaterally over the region of the auditory cortex. The voltage-sensitive dye RH-1691 (Optical Imaging, Rehovot, Israel) was applied by soaking a piece of absorbent foam in dye solution [0.5 mg/ml in artificial cerebrospinal fluid (ACSF)] and placing this over the cortical surface for 2 h, with additional dye solution dripped over the foam, 1 h into the staining. After 2 h of staining, the cortical surface was rinsed with ACSF, a melted agarose solution (1.5% in saline) was applied over the cortex, and a transparent coverslip was used to seal the craniotomy and to melt the agarose solution (1.5% in saline) was applied over the cortex, After 2 h of staining, the cortical surface was rinsed with ACSF, a

The best frequency of a pixel was then defined as the frequency eliciting the maximal response, across all frequencies and latencies. The best frequency of a pixel was held constant and the frequency of the B tone corresponding to 0, 1, 3, 9, or 18 semitones (ST; 1 ST corresponds to 1/12 octave) higher than the A tone. Both tones were presented at the same rate of 2, 4, or 8 Hz, as indicated in the text. The duration of each tone was 125 ms in the case of the 2-Hz presentation rate, 62 ms in the case of the 4-Hz rate, and 31 ms in the case of the 8-Hz rate. The sequences of tones lasted 6 s total for each streaming condition trial, with each condition repeated 10 times. Ten seconds of silence separated different trials. For the “ABA” triplet streaming paradigm, the stimulus parameters were similar, except that the A tones were presented at 4 Hz, and the B tones were presented at 2 Hz. For all variations of the auditory streaming paradigm, the A-tone frequency was chosen to be 2, 4, or 8 kHz; tonotopic cortical locations were plotted relative to this frequency to allow pooling across animals.

Frequency tuning and grouping of imaging pixels into cortical tonotopic bins. For the measurement of best-frequency tuning, data acquisition comprised ~120 total s, during which the multitone stimulus (see above) was presented. Receptive fields were derived individually for each pixel by forward correlation, i.e., calculating the average response time locked to the onset of tone pips of each frequency (Farley and Noreña 2013; Valentine and Eggermont 2004). However, their effects on best-frequency measurements and streaming paradigm responses (following trial averaging) were negligible and so, were not treated.

For the analysis of ongoing activity

J Neurophysiol • doi:10.1152/jn.00545.2015 • www.jn.org
cessed to remove baseline drift (Gaussian filter, width 1 s) and then temporally smoothed (Gaussian filter, width 20 ms). In each animal, each streaming condition was repeated 10 times, and these trials were averaged. The representations (see Fig. 2) visually illustrate the data following these processing steps and after averaging across animals.

To calculate the evoked-response amplitude elicited by each tone during the streaming paradigm for each tonotopic bin of area AI, we measured the maximum amplitude reached within a 16-ms window, centered 34 ms following the presentation of each tone (this window was empirically chosen based on the measured latency of the VSDI signal). We subtracted from this a baseline value, defined as the maximum amplitude reached in a 16-ms window centered at the time of tone presentation (i.e., before the cortical response was influenced by that tone). Thus the evoked-response amplitude reflects the short-latency change in fluorescence elicited by each tone presentation. Although evoked-response amplitudes to successive tones decreased during the 1st s of the streaming paradigm, responses achieved a steady-state level thereafter (see Fig. 3). Thus the steady-state response (for each cortical bin) to a given tone was defined as the average evoked response elicited by all presentations of that tone occurring between 1 and 6 s into the streaming paradigm (see Figs. 4A, 5B, and 6B). The absolute levels of fluorescence vary among animals, based on the efficiency of staining with the voltage-sensitive dye. Thus to permit averaging the evoked-response amplitudes across animals, it was necessary to normalize the responses in each animal. Specifically, we defined as “1” the steady-state condition in response to the A tone (see Fig. 4A) from the tonotopic bin of the cortex whose best frequency matched the A tone and normalized all responses to this value. Likewise, responses were normalized to the A-tone response in the 2-Hz condition (see Fig. 5B). Responses were normalized to the first A-tone response in the ΔF = 1 ST condition (see Fig. 6B).

Temporal coherence calculation. The algorithm for calculating temporal coherence (TC) was modified from a similar calculation previously performed on single-unit peristimulus time histograms during the streaming paradigm (Elhilali et al. 2009). This particular framework for calculating TC was chosen so that we could compare our results with those obtained by Elhilali et al. (2009). Membrane potential traces were first averaged over all pixels from each 1/4-octave tonotopic bin of the cortex, and the averaged trace was temporally filtered, as indicated above. Data from all trial repeats were averaged. TC was then calculated between the resulting membrane potential traces of each pair of tonotopic bins in the following way. First, the data traces from each bin were filtered separately at four different temporal rates of 2, 4, 8, and 16 Hz with finite impulse-response filters (1-s duration). One notes that this filtering stage differs from that used in Elhilali et al. (2009), who used a complex filter. Second, we performed a cross-correlation (or inner-product operation) between the filtered traces derived from a given pair of tonotopic bins. Finally, the mean correlation value over all four filter rates was averaged to obtain the TC value between a given pair of tonotopic bins. The TC matrices shown in the figures display this TC value calculated between all pairs (17 × 17) of tonotopic bins. We verified that the results obtained with the TC algorithm described here were similar to those obtained from using the algorithm scripts kindly provided by Elhilali et al. (2009). Eigenvalues were calculated from the resultant TC matrices of each individual animal and condition using singular value decomposition.

RESULTS

Defining the tonotopic organization of the auditory cortex. To study population neural responses during auditory streaming, we used in vivo VSDI, an optical technique that measures a membrane potential-related signal from populations of neurons in the superficial layers of the cortex at high spatial and temporal resolution (Petersen et al. 2003; Shoham et al. 1999). We sampled responses from a 25-mm² area of the cortex, which in the guinea pig allows imaging the entire tonotopic extent of area AI (Farley and Noreña 2013). We began the experiments by characterizing the acoustic best-frequency tuning of the 10,000 imaging pixels within this region (Fig. 1). The neural response was measured from each pixel to acoustic stimuli consisting of pure tones having frequencies that spanned 6 octaves (0.5–32 kHz, in steps of 1/8 octave). The temporal response profile of two example pixels in response to each of 49 test frequencies is shown (Fig. 1, A and B). The presentation of a restricted range of frequencies resulted in a short-latency increase in fluorescence, reflecting an increase in the membrane potential. By determining the frequency eliciting the maximal response for all individual pixels, the tonotopic layout of auditory cortex AI was revealed (Fig. 1C). In subsequent auditory streaming experiments, this detailed frequency-tuning information was used to describe the population pattern of responses as a function of tonotopic location, rather than spatial position, within the AI cortex.

Spatialtemporal dynamics of the cortical membrane potential during auditory streaming. We next examined the pattern of membrane potential responses across the tonotopically defined auditory cortex during the canonical auditory streaming paradigm. In the first set of experiments, we played two alternating pure-tone stimuli, A and B, each at the rate of 4 Hz. Across conditions, we parametrically varied the frequency separation between the two pure tones from 0 ST (where A and B have the same frequency) to 18 ST (corresponding to 1.5 octaves between A and B). We also included a condition where the B tone was absent (the A-only condition).

Fig. 1. Frequency tuning and tonotopic maps. A: acoustic frequency (FREQ.) receptive fields from example (2 × 2 pixel) regions of primary auditory cortex area (AI), showing the averaged temporal fluorescence response profile to each of 49 different frequencies presented during a multitone acoustic stimulation paradigm. ΔF/F, fluorescence over the baseline fluorescence value. B: image of cortical vasculature within the 5 × 5-mm imaging area from this animal, showing the locations of the 2 example pixels whose frequency receptive fields are shown. Original scale bar, 1 mm. C: each pixel is color coded according to which of the 49 frequencies elicited the maximal response. This yields the tonotopic map of area AI.
In Fig. 2, the overall spatiotemporal dynamics of the cortical responses for each streaming condition are displayed. To obtain each representation, individual pixels of the two-dimensional surface of the cortex with similar best-frequency preference were first grouped into tonotopic bins with 1/4-octave resolution (Farley and Noreña 2013). In each image, the color-coded response over time is shown for 17 adjacent tonotopic bins covering a 4-octave extent of the cortex. The center bin of each image represents the responses from those pixels whose best frequency matched the frequency of the A tone. The images represent average data collected from seven animals and 10 repeats of each streaming condition in each animal.

The presentation of the first A tone of the streaming paradigm resulted in a strong excitation across the entire 4-octave imaged region of the auditory cortex. The broad spatial distribution of responses is expected, given that the stimulus had a relatively high intensity (70 dB SPL) and because subthreshold responses have a broad distribution relative to spiking responses (Moore et al. 1999; Noreña and Eggermont 2002). Over the subsequent 1st s of each streaming paradigm condition, cortical responses to subsequent tones became increasingly selective, reflecting the operation of dynamic processes. In the case of the 0-ST condition (Fig. 2), the response amplitude diminished over time, and the area of activation became more spatially restricted around the cortical representation region corresponding to the A (and B-)tone frequency. For conditions with greater frequency separation between the A and B tones, a differential cortical response across the tonotopic axis to the A and B tones emerged over time. However, even with large separations between the A and B tones (e.g., 9 and 18 ST), a large tonotopic extent of the cortex was modulated by both tones. This indicates that even at these large-frequency separations, the cortical membrane potential responses to two successive tones were not fully spatially segregated across the cortex. Subsequent analyses explored in more detail the temporal and spatial (i.e., tonotopic) aspects of the responses, respectively.

**Temporal dynamics of the cortical membrane potential during auditory streaming.** We next examined the temporal dynamics of the neural activity patterns. Figure 3, A and B, shows the time course of membrane potential activity during the streaming paradigm for a restricted tonotopic region of the cortex, i.e., the responses from those AI pixels whose best frequency matched the A tone. In Fig. 3A, the full activity time course is shown, and in Fig. 3B, the maximal response magnitude elicited by each individual A and B tone presentation is plotted. The plots show that in all streaming paradigm conditions (0–18 ST), the response strengths change rapidly during the 1st s of the paradigm but remain at a mostly constant level thereafter.

To quantify these dynamics, we fitted the evoked-response amplitude time course data (Fig. 3B) to a double-exponential equation having the form $a \cdot e^{-b(t)} + c \cdot e^{-d(t)}$. A double exponential...
Fig. 3. Time course of the cortical membrane potential response during the auditory streaming paradigm. A: membrane potential activity, during the auditory streaming paradigm, from the cortical tonotopic bin, whose best frequency matched the frequency of the A tone. B: for the same cortical tonotopic bin, the normalized, evoked-response strength to each A and B tone repetition during the auditory streaming paradigm. C: for each tonotopic bin across a 4-octave extent of the auditory cortex, evoked-response selectivity index to the B vs. the A tone for each “AB” repetition during the streaming paradigm. A value of 1 (red) indicates response to the B tone only, and −1 (blue) indicates a response to the A tone only. D and E: points represent normalized, evoked-response strength to each A tone (D) or B tone (E) repetition during the auditory stream paradigm from the cortical bin, whose best frequency matched the A tone. Data points are replotted from B, whereas lines represent results of fitting a double exponential to the data points. For all panels, results represent the mean data over all animals. Error bars represent SE (B).
nential was chosen to capture potentially separate dynamics between the initial (i.e., 1st s) and later phases of the response time courses. The original data points along with the fitted curves are plotted in Fig. 3. D (responses to A tones) and E (responses to B tones). A number of observations are apparent from the fitted data. First, with the exception of the A-only condition, most of the response dynamics occur within the 1st s of the paradigm, after which responses remain at a steady-state level. Namely, for all streaming paradigm conditions, the amplitude of the time constant for the first exponential was ~100-fold greater than for the second exponential. Furthermore, the response strength decayed ~70% over the 1st s of the streaming paradigm, whereas only an ~10% further decrease occurred over the next 5 s. Second, although the steady-state response level differed greatly between conditions, the time constant for the initial decay was not significantly different between conditions.

The dynamics discussed so far relate to the pixels of the cortex representing the A tone. It is possible that different dynamics might be uncovered by examining other tonotopic regions of the cortex. Thus in Fig. 3C, the response dynamics are plotted for the entire 4-octave region of the cortex. For each successive “A-B” tone sequence during the streaming paradigm, we plotted for each tonotopic bin the relative response amplitude to the A tone vs. the B tone. The relative amplitudes are presented as a selectivity index (B.resp − A.resp)/(B.resp + A.resp), where 1 indicates response to the B tone only, and −1 indicates a response to the A tone only. It is clear that the response profile across the 4-octave region of the cortex, in terms of the relative response strengths to the A vs. B tones, changes during the 1st s of the paradigm but remains at a mostly constant level thereafter for all conditions.

In summary, analysis of the temporal dynamics of the cortical membrane potential indicates that responses change appreciably during the 1st s of streaming but remain at a steady-state level thereafter. Furthermore, these response dynamics are similar across streaming conditions. We next wished to characterize the detailed spatiotemporal properties of the cortical membrane potential during this steady state.

Cortical membrane potential profile during steady state: effect of frequency separation. Having identified the time at which responses achieve a steady-state level during the auditory streaming paradigm, i.e., ~1 s after the beginning of the sequence, we next asked in more detail what the spatial (tonotopic) and spatiotemporal profiles of the cortical membrane potential responses during this steady-state period are. We asked whether changes across conditions in the spatial pattern of the cortical membrane potential responses correlate with known perceptual trends, as measured in human subjects (see Discussion), for perceiving the tone sequences as segregated or integrated. The best-characterized effects on perception are those of changing the frequency separation (ΔF) between the tones (Michel et al. 2005; van Noorden 1977).

In Fig. 4A, the steady-state tonotopic profile of evoked cortical responses during the streaming paradigm is illustrated. In this figure, we plot the amplitude of the response evoked by the A or B tone relative to the background level (immediately before the given tone presentation). When ΔF was 1 ST or greater, the A and B tones elicited distinct tonotopic activation profiles (Fig. 4A). As ΔF increased across conditions, lower-frequency representation regions responded more selectively to the A tone, whereas higher-frequency representation regions respond more selectively to the B tone. However, an important feature of the membrane potential responses was a large (>2 octaves) overlapping cortical region—activated by both the A and B tones—which was present in all of the ΔF conditions tested in this experiment (Fig. 4A). Overall, these data show that as the frequency separation between two alternating rhythmic tones increases, topographically separated regions respond more selectively to one input stream over the other. Furthermore, the magnitude of the response to one input stream increases relative to the response to both streams. However, a portion of the cortex maintains the potential to respond to both tones across conditions; we refer to this region responding to both tones as the “integrated representation region.”

Recent studies have suggested that the tendency to hear two alternating tones as an integrated or segregated percept may depend on the temporal coherence (TC) of neural activity across the auditory cortex (Elhilali et al. 2009). This idea is complementary to the spatial segregation hypothesis, by considering more explicitly the temporal dimension of neural activity in addition to the spatial or tonotopic dimension. The population recordings performed here made it possible to measure directly the TC of cortical membrane potential responses during auditory streaming. In Fig. 4B, the value of TC between each pair of tonotopic bins across a 4-octave extent of the auditory cortex is illustrated. In the 0-ST condition, one tonotopic region is visible with high TC relative to surrounding regions. As the frequency separation between the A and B tones increased (from 0 to 18 ST), two increasingly distinct regions with higher internal coherence and lower between-region coherence emerged. When the B tone was removed entirely, only a single region of high internal TC was visible. Figure 4C shows the TC matrices from the first and second eigenvalues following singular value decomposition analysis. The matrix from the first eigenvalue in all cases corresponded to a single region of high TC. The matrix from the second eigenvalue appeared as two regions of high internal coherence but with low coherence between them. As ΔF increased across conditions, the strength of the second eigenvalue increased relative to that of the first (Fig. 4C; see also Fig. 7). These data thus indicate that the TC across the cortex decreases for the cortical membrane potential as ΔF increases in the auditory streaming paradigm, as has been demonstrated previously for suprathreshold activity (Elhilali et al. 2009).

Cortical membrane potential pattern during steady state: effect of presentation rate. In the auditory streaming paradigm, changes in parameters besides ΔF are known to affect perceptual outcomes. For example, at a given ΔF, increasing the rate of tone presentation raises the probability of hearing the two tone sequences as segregated (Bregman 1990). We examined how changing the rate of tone presentation (while maintaining ΔF constant at 9 ST or 1/4 octave) would affect the pattern of cortical membrane potential responses. Figure 5A illustrates that the spatiotemporal dynamics of the cortical membrane potential were influenced greatly by changes in the rate of tone presentation. Similarly, the steady-state tonotopic response profiles were largely different for different rates of tone presentation (Fig. 5B). Notably, for the 8-Hz tone-presentation rate, the cortical tonotopic profiles of the evoked responses elicited by the A and B tones had no overlap. The cortical TC
patterns also changed dramatically with changes in tone-presentation rate. Whereas one tonotopic region with high TC, relative to surrounding regions, was present in the 2-Hz condition, two regions with high internal coherence but low coherence between each other dominated in the 8-Hz condition (Fig. 5, C and D). These results further support the idea that the tendency to segregate streams perceptually is accompanied by decreasing spatial overlap between evoked cortical responses and decreasing TC across the cortex.

Cortical membrane potential pattern during steady state: ABA “gallop” paradigm. So far, this study has examined the ABAB auditory streaming paradigm, where two alternating tones are played at identical rates. In the ABA gallop variant of the paradigm, one tone sequence is presented at twice the rate as the other (Fig. 6A). Despite the different temporal acoustic sequences present between the AAB and ABA paradigms, similar perceptual phenomena are observed; namely, the increase of ΔF between two alternating tones raises the probability of hearing them as perceptually distinct sequences (Carllyon et al. 2001; Micheyl et al. 2005; van Noorden 1977).

We measured the spatiotemporal pattern of cortical membrane potential responses for three conditions of the ABA paradigm, differing in their ΔF values (Fig. 6A). The tonotopic pattern of cortical membrane potential responses observed
during the “steady-state” portion of the ABA paradigm is plotted in Fig. 6B. Responses to the first and second A tones of the triplets are plotted separately, since their contexts within this paradigm are unique. In the 1-ST condition, where the dominant percept (measured in human subjects) is that of a single “triplet” stimulus (Carlyon et al. 2001; Micheyl et al. 2005; van Noorden 1977), the evoked responses within all tonotopic cortical regions were dominated by the first A tone of the triplet compared with the remaining (B and second A) tones of the triplet. As ΔF increased, the responses to the B tone and
to the second A tone increased relative to the response to the first A tone. In the 9-ST condition, cortical regions above the B-tone frequency responded more strongly to the B tone, whereas regions below the A-tone frequency respond more strongly (and equally) to the two A tones. Interestingly, the dominant percept in the 9-ST condition is of two streams (as assessed in human subjects), with the lower frequency stream occurring at twice the rate of the higher frequency stream (Carlyon et al. 2001; Micheyl et al. 2005; van Noorden 1977). The changes in the TC pattern with ΔF in the ABA gallop paradigm were similar to those observed in the ABAB paradigm, despite the different stimulus temporal frequencies present between the two paradigms. Namely, one region of high TC was present in the 1-ST condition, whereas a second
region of high coherence, centered around the frequency of the B tone, became increasingly apparent as $\Delta F$ increased (Fig. 6, C and D).

**Relationship between TC and known perceptual trends in all three streaming paradigm variations.** The results so far suggest that the tonotopic pattern of the cortical membrane potential response and its TC across the cortex changes systematically with parameters known to affect perceptual outcomes in several versions of the auditory streaming paradigm. We wished next to compare directly and quantitatively the behavior of the cortical membrane potential across all paradigms and conditions tested in the study. We focused on the TC measure, given that it provides the most complete (tonotopic as well as temporal) description of the neural activity. The metric we chose is based on the relative strength of the second vs. first eigenvalue of the TC matrices, which can be estimated using singular value decomposition (Elhilali et al. 2009). In the ABAB streaming paradigm, we found that the value of this metric increases steadily with $\Delta F$ (Fig. 7), indicating that the metric correlates with the increased perceptual tendency to hear two streams as $\Delta F$ increases. This result also held true when testing the ABA paradigm at three values of $\Delta F$. Finally, the same metric was also highly sensitive to changes in the rate of stimulus presentation, with increasing rates leading to higher values of the metric. Together, the results show that decreased TC of subthreshold activity across the tonotopic axis of the cortex, whether resulting from changes to tonotopic or temporal stimulus parameters, correlates with the perceptual trend for hearing the tone sequences as two streams.

**DISCUSSION**

The auditory streaming paradigm is a simplified model of the cocktail party problem when one sound source has to be segregated from an acoustic mixture (Bregman 1990; Cherry 1953). This paradigm, which has been well characterized psychophysically (Moore and Gockel 2012), is also used for examining how patterns of neural activity relate to perceptual trends for the problem of sequential grouping. Whereas auditory cortical activity has been measured previously during streaming paradigms, the current study is the first to do so using VSDI. The main advantage of VSDI over previous electrophysiological approaches is that it measures neural activity across a large extent of the auditory cortex at high spatial and temporal resolution (Grinvald and Hildesheim 2004). Therefore, this method is well adapted to test some of the hypotheses advanced in the literature relative to the mechanisms of sequential grouping, i.e., whether sequential grouping is related to the spatial separation and/or TC between neural populations (Elhilali et al. 2009; Fishman et al. 2001; Hartmann and Johnson 1991; Micheyl et al. 2005). Furthermore, VSDI provides the capability to measure subthreshold responses across the cortex, in contrast to imaging with calcium sensors. Patterns of subthreshold excitability have recently been proposed to play a central role in organizing sensory representations (Lakatos et al. 2008) but have not yet been examined during the auditory streaming paradigm. Some trends observed here for subthreshold neural activity are consistent with those observed from previous studies of suprathereshold activity (Bee and Klump 2004, 2005; Bee et al. 2010; Fishman et al. 2001, 2004; Micheyl et al. 2005). However, some new findings also emerge, suggesting an original hypothesis for how subthreshold neural representations may contribute to explaining alternating perceptual outcomes during ambiguous acoustic scenes, i.e., bistable percepts.

**Spatiotemporal profile of cortical membrane potential during streaming.** Psychoacoustic studies of the auditory streaming paradigm have informed some of the key hypotheses regarding the mechanisms of sequential integration and segregation. The probability of hearing two streams in an ABA tone sequence was shown to depend on the frequency separation between the A and B tones (Moore and Gockel 2012; van Noorden 1977). This observation led to a “peripheral channeling” hypothesis, whereby auditory streaming was proposed to depend partially on the degree of spatial overlap between cochlear excitation patterns evoked by successive sounds (Beauvois and Meddis 1996; Hartmann and Johnson 1991; Rose and Moore 2000). However, auditory streams can also form when two sounds result in cochlear excitation patterns that overlap, even completely. Under these conditions, segregation could not depend on peripheral channeling alone but is instead proposed to arise from differences in the temporal envelope, phase spectrum, or fundamental frequency of the sounds (Grimault et al. 2002; Roberts et al. 2002; Vliegen and Oxenham 1999). It remains possible, however, that a central “channeling” mechanism could contribute in those cases where neurons sensitive to higher-level acoustic features, such as pitch (Bendor and Wang 2005; Griffiths and Hall 2012), exist and are spatially segregated. Some additional key psychoacoustic observations during auditory streaming are that segregation builds up over time and is sensitive to the stimulus presentation rate (Bregman et al. 2000). These observations can also be explained in the context of the channeling mechanism described above. Repeated acoustic stimulation leads to a reduction of neural activity, which is more rapid and pronounced as the rate of stimulus presentation increases (Kvale and Schreiner 2004; Ulanovsky et al. 2004; Wehr and Zador 2005). Stronger reduction of neural activity would result in decreased spatial overlap between neural populations activated by successive sounds and therefore would contribute to stream segregation.
Our observations on cortical membrane potential are consistent with the hypothesis that the degree of spatial overlap between neural populations, excited by successive sounds, is a key determinant of perceptual segregation and are also in broad agreement with previous suprathreshold data. First, as the frequency separation between two alternating tones increases, cortical membrane potential responses become increasingly selective to one of the two tones. These results are qualitatively consistent with those observed previously in single- and multiunit studies (Fishman et al. 2004; Micheyl et al. 2005). Second, the cortical membrane potential responses become increasingly selective to one of the two tones as the presentation rate used in the acoustic sequence increases. These results are consistent with those reported by Fishman et al. (2001), who found that as the tone-presentation rate was increased during the streaming paradigm, response suppression for multiunit activity was observed preferentially for nonbest-frequency tones, such that units became more selective for their preferred frequency. Third, the spatial separation between cortical regions activated by successive sounds is enhanced over time, namely over the 1st s of the streaming paradigm (Fig. 3C). This is in broad agreement with suprathreshold recordings performed in macaque AI and in the avian homologue of AI during the ABA sequence version of the streaming paradigm, where response selectivity and by extension, spatial segregation increase over time (Bee et al. 2010; Micheyl et al. 2005). Overall, our results are consistent with earlier studies of suprathreshold activity, in suggesting that the degree of spatial overlap decreases over time and with acoustic manipulations (∆F and presentation rate) known to increase the tendency for perceptual segregation.

More recently, the concept of TC between neural populations has been proposed as an alternative to the spatial channeling hypothesis as a mechanism underlying auditory stream segregation (Bee et al. 2010; Elhilali et al. 2009). It is known that sounds with large frequency difference presented synchronously can be fused into a single object or stream even when they result in neural activity in spatially segregated neural populations (Carlyn 2004; Darwin 1997; Elhilali et al. 2009). The TC of neural activity, being sensitive to the temporal relationship between sounds, appears to be able to account for this observation (Elhilali et al. 2009). The membrane potential responses obtained in this study were analyzed within the framework developed by Elhilali et al. (2009). This framework was chosen to be able to compare our results directly with those of Elhilali et al. (2009), who measured spiking responses. We calculated TC across the cortex from the cortical membrane potential signal measured during the auditory streaming paradigm. When the TC matrices calculated across multiple frequency bins were decomposed into principal components, it was observed that the first component corresponded to a single cortical region of high coherence, whereas the second component corresponded to two regions of high internal coherence but with low between-region coherence. The first and second components can be interpreted as one- vs. two-stream representations in the cortex, respectively. In the case where the percept is unambiguously that of one stream (the 0-ST case or the A-only case), the first eigenvalue (i.e., one-stream representation) dominated in the cortical membrane potential responses. However, as the frequency difference between the A and B tones in the streaming paradigm increased, the relative strength of the second eigenvalue (i.e., the two-stream representation) increased. These results are consistent with the study of Elhilali et al. (2009), who showed that suprathreshold responses have decreased TC across the cortex with increasing ∆F. Additionally, our results extend the findings of Elhilali et al. (2009) in several ways. First, our study shows for the first time a complete view of the spatiotemporal activity in the auditory cortex during the auditory streaming paradigm (as opposed to the partial view provided by discrete microelectrode recordings). Second, we show that the dependence of spatiotemporal separation of neural population activity on ∆F in the streaming paradigm is present in the subthreshold cortical membrane potential signal. Third, we show that spatiotemporal separation also depends on tone-presentation rate (while frequency parameters are held constant). We also note differences between the behavior of the subthreshold activity (as measured here) and suprathreshold activity (as measured previously), as will be discussed below.

Whereas the conditions under which stream segregation is expected to occur may differ across species, the trend for an increasing probability of stream segregation with increasing ∆F and presentation rate is expected to hold across species. Thus our comparison of membrane potential patterns across streaming paradigm conditions, where acoustic parameters were systematically varied, supports the hypothesis that the TC of cortical activity contributes to stream segregation.

Spatiotemporal overlap of activity and perceptual ambiguity. Whereas our results are in general agreement with previous studies measuring suprathreshold signals during auditory streaming (see above), they also diverge in some ways. Most importantly, subthreshold cortical membrane potential responses to individual tones, presented during the streaming paradigm, were much broader (across the tonotopic axis) than those observed previously from suprathreshold measurements under similar conditions. For example, Micheyl et al. (2005) report a very small suprathreshold response to a tone that is 9 ST away from the neural best frequency, whereas a strong cortical membrane potential response to this tone was observed in the present study (see Fig. 4A). Likewise, the TC across the cortex for cortical membrane potential responses measured in the present study (see Fig. 4B) was higher than that derived from suprathreshold activity under similar conditions (Elhilali et al. 2009).

Whereas it cannot be ruled out that anesthesia or pharmacological effects of the voltage-sensitive dye (Grandy et al. 2012) contribute to the observed differences between the subthreshold and suprathreshold responses, such differences are also predicted by the different spatial summation properties of the two signals—subthreshold responses being broader in response to simple stimuli (Carandini and Ferster 2000; Chaderton et al. 2009; Galvan et al. 2002; Moore et al. 1999; Noreña and Eggermont 2002). Furthermore, the broad cortical membrane potential responses observed here cannot be explained by a lack of spatial resolution, given that under certain conditions—namely, the ST 9, 8-Hz condition of the ABAB paradigm (cf. Fig. 5)—two regions <1 octave away had completely noncorrelated responses.

The broad cortical membrane potential responses observed here imply that under most streaming paradigm conditions tested, responses to alternating tones are not completely spa-
tio-temporally segregated. Rather, a cortical region that responds robustly to tones of both sequences (an integrated representation region) coexists with peripheral cortical regions that respond more selectively to one tone sequence ( segregated representation regions). We propose that these coexisting subthreshold membrane potential representation types may play a role in the context of auditory scene analysis, relating to the perceptual phenomena of stream bistability and resetting. Namely, perception can be unstable during the auditory streaming paradigm, alternating spontaneously between integrated (one-stream) and segregated (two-stream) percepts, whereas the acoustic environment remains constant (Denham and Winkler 2006; Pressnitzer and Hupe 2006; Schwartz et al. 2012). Moreover, perception can be reset to the one-stream percept by an abrupt change in the acoustic sequence (Cusack et al. 2004; Roberts et al. 2008; Rogers and Bregman 1998). Attention has also been shown to influence the buildup of streaming and to bias which percept is experienced at any given time (Carlyon et al. 2001; Cusack et al. 2004; Thompson et al. 2011; van Noorden 1977).

The neural mechanisms allowing for alternative perceptual interpretations during the streaming paradigm, despite the fact that the acoustic sequence remains stable over time, remain unknown. We speculate that auditory cortical membrane potential representations could contribute to this process. According to this view, the integrated and segregated subthreshold membrane potential representations can coexist during the streaming paradigm, and each representation corresponds to a potential state (Mill et al. 2013). On the other hand, the moment-by-moment perception would correspond to a “choice” between one potential state and the other. This choice would be influenced by attention and other factors, which are known to modulate cortical activity patterns (Fritz et al. 2003; Lakatos et al. 2008; Roberts et al. 2007; Roelfsema et al. 1998; Womelsdorf et al. 2007). A key feature of this hypothesis is the persistence of the integrated and segregated perceptual outcomes. The view of multiple coexisting potential representations in subthreshold signals is consistent with data recorded from humans or nonhuman primates during auditory selective-attention experiments, where representations exist for both attended and unattended objects, the difference being that the one for the currently attended object is enhanced or phase shifted (Lakatos et al. 2013; Mesgarani and Chang 2012; Zion Golumbic et al. 2013). Our observations during the auditory streaming paradigm suggest that such a process may also be at play when the problem is whether to integrate vs. segregate sounds from an acoustic mixture.

ACKNOWLEDGMENTS

The authors thank Shihab Shamma for valuable comments on a previous version of the manuscript and Mounya Elhilali for sharing her MATLAB Scripts.

GRANTS

Support for this work was provided by the Tinnitus Research Initiative and L’Agence Nationale de la Recherche (ANR; Grant ANR-2010-JCJC-1409-1).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.
Membrane Potential Dynamics During Auditory Streaming


