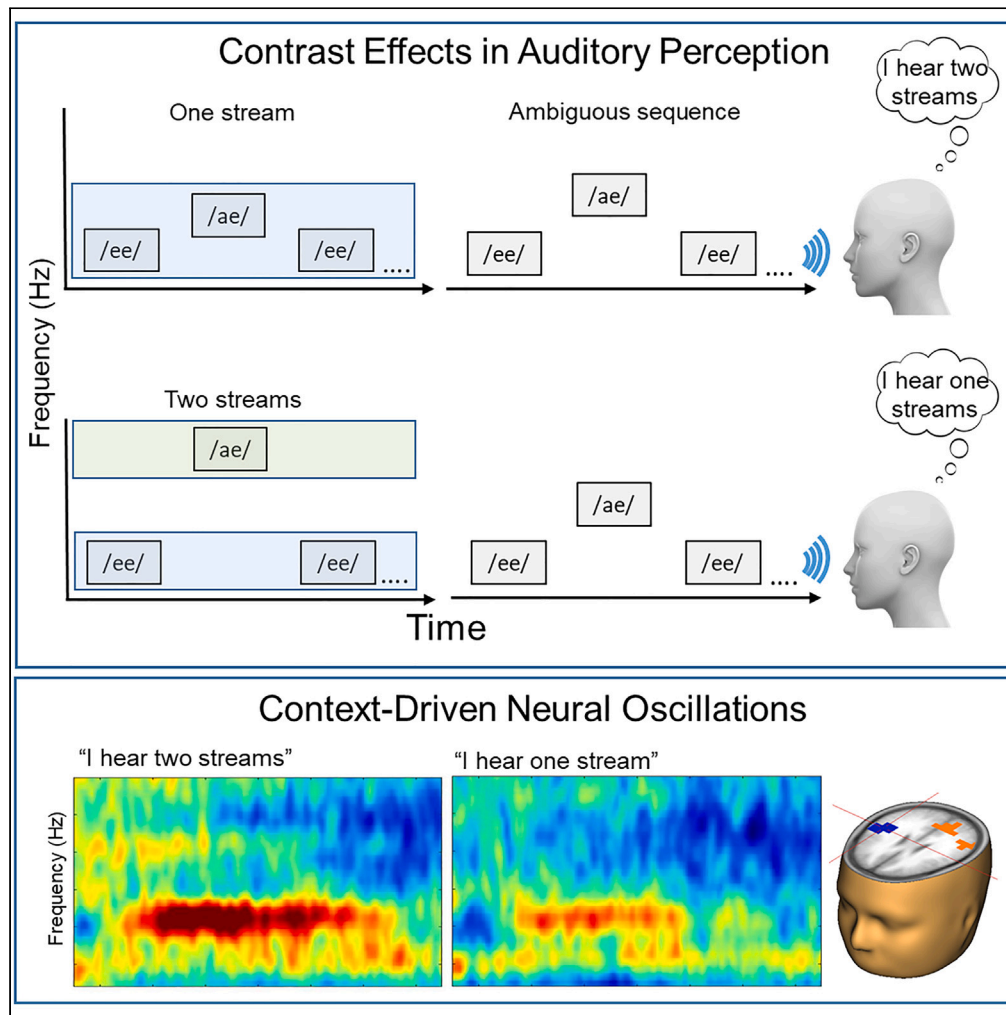


Article

Neural alpha oscillations index context-driven perception of ambiguous vowel sequences



Claude Alain,
Katharina Göke,
Dawei Shen, Gavin
M. Bidelman, Lori
J. Bernstein, Joel S.
Snyder

calain@research.baycrest.org

Highlights

Perception of an ambiguous auditory stimulus is influenced by prior context

Listeners report hearing the opposite percept of the one heard immediately before

Prior context modulates stimulus-evoked alpha power in angular and supramarginal gyrus

Alpha Power correlates with listeners' perceptual construal level of ambiguous stimuli

Alain et al., iScience 26, 108457
December 15, 2023 © 2023 The Author(s).
<https://doi.org/10.1016/j.isci.2023.108457>



Article

Neural alpha oscillations index context-driven perception of ambiguous vowel sequences

Claude Alain,^{1,2,6,*} Katharina Göke,¹ Dawei Shen,¹ Gavin M. Bidelman,³ Lori J. Bernstein,⁴ and Joel S. Snyder⁵

SUMMARY

Perception of bistable stimuli is influenced by prior context. In some cases, the interpretation matches with how the preceding stimulus was perceived; in others, it tends to be the opposite of the previous stimulus percept. We measured high-density electroencephalography (EEG) while participants were presented with a sequence of vowels that varied in formant transition, promoting the perception of one or two auditory streams followed by an ambiguous bistable sequence. For the bistable sequence, participants were more likely to report hearing the opposite percept of the one heard immediately before. This auditory contrast effect coincided with changes in alpha power localized in the left angular gyrus and left sensorimotor and right sensorimotor/supramarginal areas. The latter correlated with participants' perception. These results suggest that the contrast effect for a bistable sequence of vowels may be related to neural adaptation in posterior auditory areas, which influences participants' perceptual construal level of ambiguous stimuli.

INTRODUCTION

Our visual and auditory environment comprises complex and sometimes ambiguous stimuli that elicit multiple interpretations, which have perceptual consequences. In the auditory domain, bistable perception is often illustrated using the ABA paradigm. Participants are presented with a sequence of repeating ABA patterns comprising low-frequency (A) and high-frequency tones (B). When the frequency separation between tones A and B is small, participants report hearing a "galloping" rhythm. When the frequency separation between the two tones is large, participants report hearing two concurrent and separate streams of sounds. At an intermediate level of frequency separation, participants experience bistable perception, where their percept alternates back and forth between the perception of one and two streams.^{1,2} This auditory bistability is associated with neural activity in a widely distributed neural network including inferior colliculus,³ thalamus,^{4,5} auditory cortex,^{4,5} and parietal cortex.^{6,7}

Perception of ambiguous (e.g., bistable, reversible) stimuli is influenced by preceding context or priming stimuli. In some cases, the perceptual interpretation of a visual ambiguous stimulus matches that of the preceding context, which relates to both bottom-up (e.g., stimulus-driven) and top down factors (e.g., attention).^{8,9} In others, the observer perceives the alternate of the preceding stimulus (i.e., prior context).¹⁰ This is referred to as the contrast effect, a magnification of the difference between two successive stimuli. Such contrast effects have also been observed with bistable auditory stimuli.^{11–14} In these studies, participants were presented with a contextual cue that comprised a repeating ABA-pattern. The physical difference between the A and B tones was manipulated to induce the perception of one or two streams. An ambiguous sequence followed the contextual cue. Participants reported hearing the ambiguous sequence as one stream more often when the contextual cue promoted the perception of two streams and vice versa. This contrast effect depended on stimulus-related factors. That is, prior perception of a contextual cue sharing the same stimuli as the bistable stimuli did not yield a contrast effect. If anything, participants had a tendency to report the same percept as the one from the contextual cue.¹¹ Thus, the context effect on perceptual reversals appears to depend on stimulus-specific context rather than prior perception. This is consistent with the importance of neural adaptation in general models of bistable perception¹⁵ and more specifically models of auditory stream segregation.^{16,17}

In a prior study,¹⁴ we investigated the neural correlates of speech segregation and bistable perception by combining scalp recording of event-related potentials (ERPs) with a paradigm analogous to Snyder et al.¹² Participants were presented with a contextual sequence and an ambiguous ABA-sequence of vowels (/ee/-/ae/-/ee/). In the contextual sequence, the first formant difference (Δf_1) between the two vowels was either small, intermediate, or large. In the ambiguous sequence that followed, the difference in the first formant between the two vowels

¹Rotman Research Institute, Toronto, ON M6A 2E1, Canada²Department of Psychology, University of Toronto, Toronto, ON M5S 3G3, Canada³Department of Speech, Language and Hearing Sciences and Program in Neuroscience, Indiana University, Bloomington, IN 47408, USA⁴Department of Psychiatry, University of Toronto and University Health Network, Toronto, ON M5G 2C4, Canada⁵Department of Psychology, University of Nevada, Las Vegas, NV 89154, USA⁶Lead contact

*Correspondence: calain@research.baycrest.org

<https://doi.org/10.1016/j.isci.2023.108457>

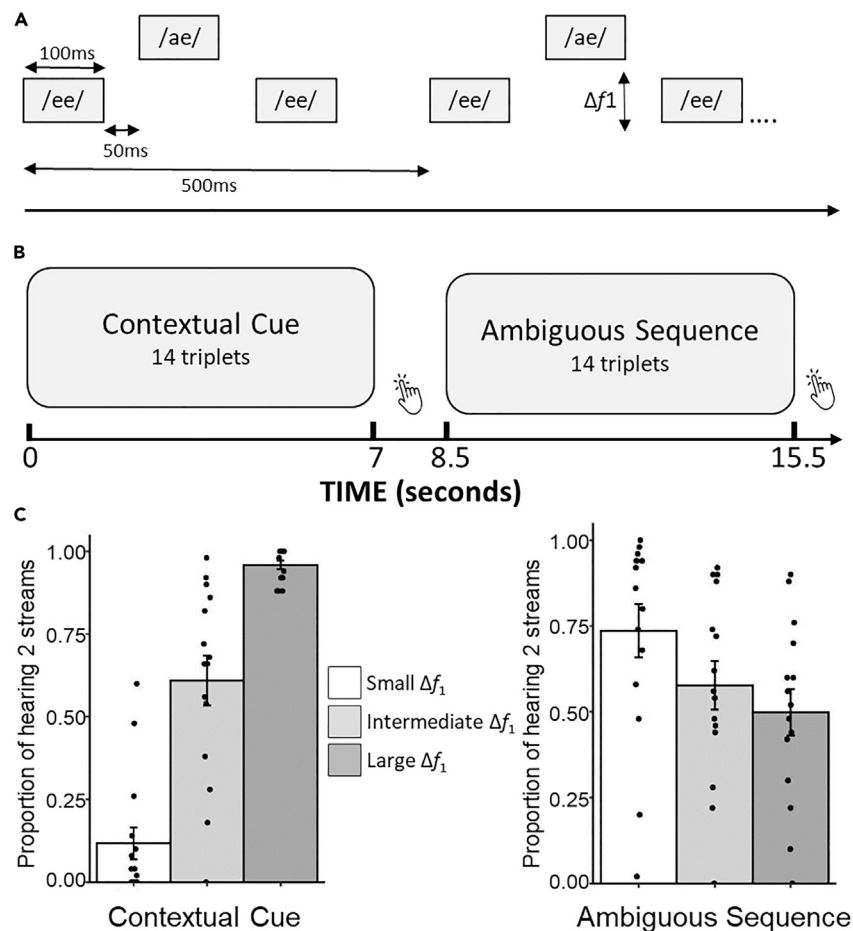


Figure 1. The speech ABA paradigm and behavioral performance

(A) Graphical depiction of two vowel triplets. Each triplet lasted 400 ms and contained three vowels. The interval between vowels was 50 ms, and the interval between triplets was 100 ms.

(B) Schematic of a trial. Each trial consisted of a contextual cue comprising 14 vowel triplets followed by an ambiguous sequence also comprising 14 triplets. Participants responded immediately after the sequence indicating whether they heard one or two streams.

(C) Group mean proportion (percent) of trials heard as two streams as a function of sequence type and first formant difference. Error bars = ± 1 standard error of the mean. Dots = individual data.

was always intermediate. The likelihood of reporting hearing two streams increased with Δf_1 and coincided with changes in ERP amplitude between 100 and 200 ms after the /ae/ vowel onset, which was localized in auditory cortices. During the ambiguous sequence, we observed a contrast effect such that participants' perception of the ambiguous sequence was opposite to that of the preceding stimulus. This contrast effect coincided with a difference in ERP amplitude, which was characterized by an increased positivity between 170 and 300 ms after the first vowel of the triplet (i.e., 20–150 ms after the /ae/ vowel). Distributed source analysis of this short transient change in ERP amplitude was localized to auditory cortices, with an additional source in the left prefrontal cortex. The prior ERP analyses have limitations in that they did not consider non-phase locked activities that may be sensitive to the perception of a bistable stimulus.^{18–20} Time-frequency analyses can better characterize the temporal dynamics of bistable perception and shed light on the neural networks associated with bistable perception. The present study aims to extend our prior findings¹⁴ by examining evoked and induced neural activity for the whole duration of the ambiguous sequence using time-frequency analysis.

Studies on bistable auditory stimuli have not yet resolved at which stage in the hierarchy of auditory and supramodal areas, neural activity is associated with participants' tendency to respond opposite to how they responded to a previous stimulus. Evidence from electrophysiological studies suggests that perceptual switching between the two possible interpretations is associated with changes in oscillatory brain activity in alpha and beta bands,^{18,20–22} which have been associated with activity in parietal areas and sensorimotor regions.^{23–25} In this study, we re-analyzed the EEG data collected in our previously published study of speech segregation,¹⁴ and compared oscillatory activity elicited by an ambiguous auditory stimulus preceded by a contextual cue (Figures 1A and 1B). We hypothesized that the contrast effect is related to changes in alpha and beta power elicited by the bistable stimulus.

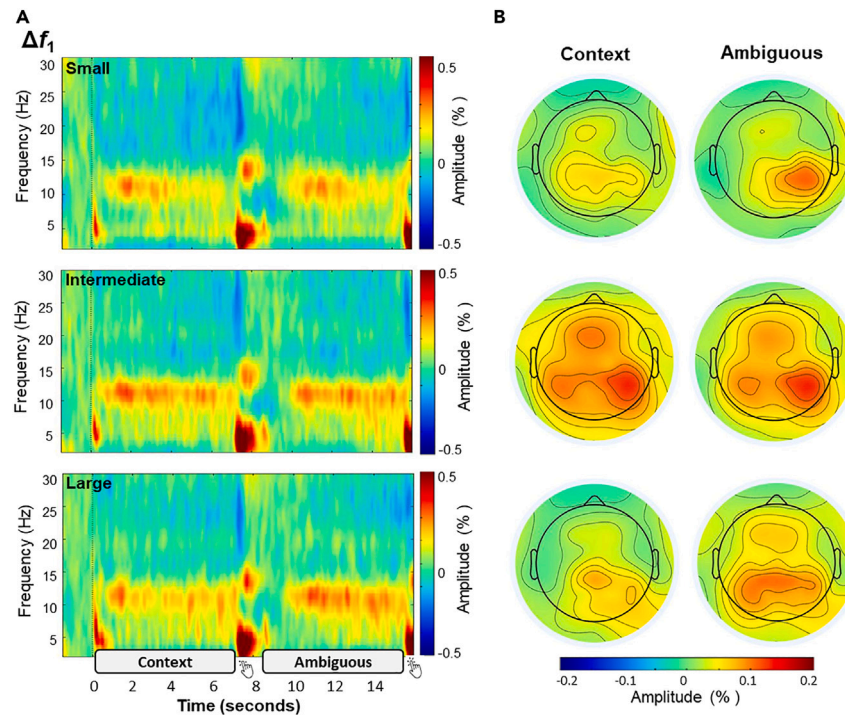


Figure 2. Steady-state evoked neural oscillations and distribution of alpha power

(A) Group mean temporal spectral evolution (TSE) time-locked on the onset of the contextual cue when the difference between the first formant (Δf_1) was small, intermediate, or large. The spectrograms show oscillatory activity from the midline central-parietal electrode (CPz). (B) Isocontour maps for the mean alpha power (8–13 Hz) between 1–7 s and 9.5–15.5 s, representing responses during the contextual cue vs. ambiguous sequence, respectively.

RESULTS

Behavioral data

Figure 1C shows the group mean proportion of trials in which participants report hearing two streams as a function of Δf_1 . For the context sequence, participants' likelihood of reporting hearing two auditory streams increased with increasing Δf_1 ($F(2,28) = 68.914$, $p < 0.001$, $\eta_p^2 = 0.831$, all pairwise comparisons $p < 0.001$). This pattern was reversed for the ambiguous sequence, with participants being less likely to report hearing two streams when the ambiguous sequence was preceded by large Δf_1 , and conversely, were more likely to report hearing two streams when the ambiguous (i.e., bistable) sequence was preceded by a context sequence with a small Δf_1 ($F(2,28) = 18.597$, $p < 0.001$, $\eta_p^2 = 0.571$, all pairwise comparisons $p < 0.05$). There was no difference in perception of the intermediate Δf_1 as a function of sequence type (contextual sequence versus ambiguous sequence, $t(14) = 1.438$, $p = 0.172$). These results demonstrate a contrast effect in that observers' perceptual interpretation of the ambiguous sequence was more likely to be opposite of their interpretation of the contextual cue.

Prior research has shown that stream segregation is not instantaneous but builds over time.^{2,26,27} Therefore, the percept of two concurrent streams may have been weaker than that of one stream, yielding a smaller contrast effect when the contextual sequence comprised large Δf_1 . We tested whether the switch in perception varied as a function of the Δf_1 . For each participant, we calculated the proportion of changes in perception between the context and the ambiguous sequence. A paired t test showed that the contrast effect was comparable when the contextual sequence comprised a small or large Δf_1 ($t(14) = 0.89$, $p = 0.39$).

Electrophysiological data

Figure 2 shows the group mean temporal spectral evolution (TSE) spectrograms for the whole trial time-locked to the onset of the contextual cue. The contextual cue and the ambiguous stimulus sequences were associated with transient theta power (~ 5 Hz) at sequence onset, followed by sustained alpha power (~ 10 Hz) during the steady-state portion of the contextual cue and the ambiguous stimulus. Alpha synchronization was largest over the left and right parietal scalp areas. The motor response (i.e., button presses) at the end of both sequences showed transient synchronization of delta, theta, and beta power.

Participants' behavioral reports for the intermediate Δf_1 did not differ for the contextual cue and the ambiguous sequence. Therefore, the following analyses focus on the oscillatory activity elicited during the ambiguous sequences when preceded by the contextual cue with a small or a large Δf_1 . This comparison is best suited to identify the neural correlates of the contrast effect associated with prior perceptual context because these conditions modulated listeners' streaming reports.

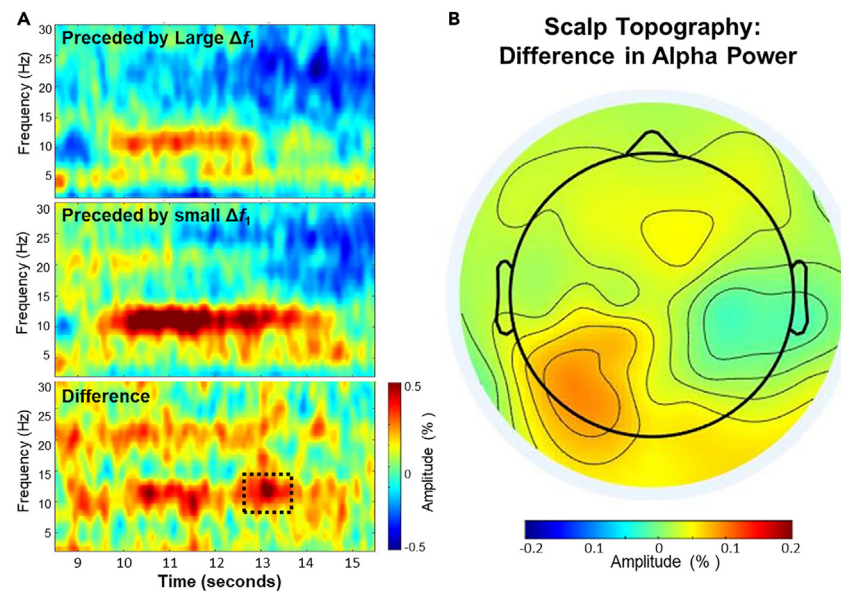


Figure 3. Context-driven neural oscillations and distribution of alpha power

(A) Group mean TSE time-locked to the onset of the ambiguous sequence preceded by large (top) and small (middle) Δf_1 , and the corresponding difference (bottom). The spectrograms show oscillatory activity from the left parietal electrode (P5). The rectangle highlights the maximum difference in alpha power. (B) Isocontour maps showing the mean alpha power (8–13 Hz) distribution for the difference between larger and small Δf_1 during the 12.6–13.6 s interval.

The cluster analysis procedure and permutation-based statistics revealed one significant spatiotemporal cluster ($p = 0.002$) (Figure 3). The cluster started at 10.1 s and ended at 15.5 s and encompassing theta, alpha, and beta bands. It peaked at 13.15 s and was maximum at 12 Hz over the left temporal-parietal, central-parietal, parietal, and parietal-occipital scalp area (TP7, CP5, CP1, P7, P5, P3, P1, PO3, and POz) and for that interval, extended to right frontal scalp area (AF4, AF8, F4, F6, F8).

We performed source reconstruction of the time-frequency data using a multiple source beamformer (MSBF) method²⁸ as implemented in BESA Research 7.1. The MSBF captures both evoked and induced oscillatory activity in a specified time-frequency range and allows for separation of bilateral activity in the two hemispheres that is often highly correlated and thus blurred in traditional source reconstruction approaches. For the small and large Δf_1 , we estimated source activity for frequencies ranging from 8 to 13 Hz and for a 1000 ms interval centered on the peak latency of the difference between the two conditions (i.e., 13100 ms, Figure 3) using $-1000 \text{ ms} - 0 \text{ ms}$ as the baseline. Using an $\alpha = 0.01$ for cluster buildup, the permutation t tests revealed three brain areas driving the perception-related differences in streaming (Figure 4). The first cluster peaked in the left angular gyrus (Talairach coordinates: $x = -59$, $y = -53$, $z = 19$), and encompassed the superior temporal gyrus and inferior parietal cortex. The second and third cluster peaked in the left ($x = -23$, $y = -26$, $z = 55$) and right primary sensory-motor area ($x = 50$, $y = -17$, $z = 28$). The latter encompassed the supramarginal gyrus.

Brain behavior correlations

We examined whether the effect of contextual cue on source activity during the ambiguous sequence related to the contrast effect observed behaviorally. Pearson correlations between the contrast effect and the difference in source activity for all three clustered identified in the MSBF analysis is presented in Table 1. The source activity in the right sensorimotor/supramarginal areas correlated with the contrast effect observed behaviorally. That is, the greater the contrast effect, the larger the difference in source activity in the right sensorimotor and supramarginal area. Additionally, the source activities in the left angular gyrus were positively correlated with source activities in left and right sensorimotor areas.

DISCUSSION

Our results offer important insights into auditory bistable perception. We show that the contrast effect coincides with changes in alpha power localized in the left angular gyrus, left and right sensorimotor areas. Importantly, the alpha power source activity in right sensorimotor areas was positively correlated with the contrast effect. These findings extend our earlier study showing Δf_1 -related changes in auditory cortices elicited by triplet onset. These findings also extend previous electrophysiological studies by revealing differences in brain activity in attention networks, including associative auditory cortices, and parietal and sensorimotor areas.

The contextual and the ambiguous sequences were both associated with transient theta power elicited by sequence onset and sustained alpha synchrony throughout the steady-state portion of the sequence. Notably, the contextual sequence influenced alpha synchrony elicited

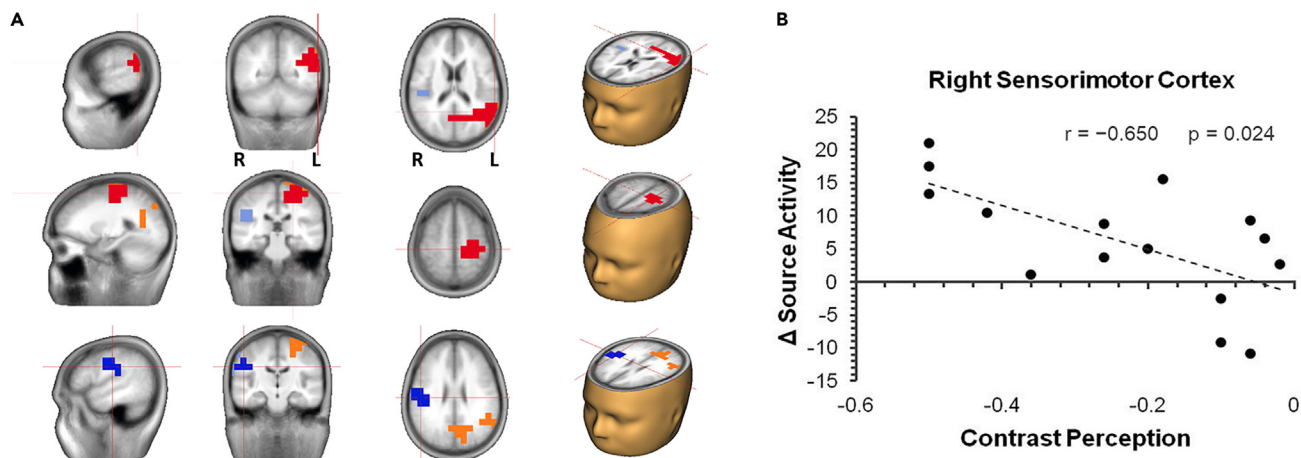


Figure 4. Source localization of alpha power

(A) Group mean difference in beamformer source alpha power (8–13Hz) during the 12600–13600 ms interval. The contrast yielded three different clusters. The first cluster was localized in the left angular gyrus (top row, red). The second and third clusters were localized in left (middle row, red) and right (bottom row, blue) primary sensory-motor areas, respectively. The red color indicates greater power (i.e., alpha synchronization) when the context was two streams rather than one. The blue color indicates lower power (alpha suppression) when the context was two streams rather than one.

(B) Scatterplot showing the relationship between the source activity in the right sensorimotor area and the contrast effect.

by the bistable/ambiguous stimulus, with more substantial alpha power over the right parietal region when the context comprised small Δf_1 (participants reported hearing one stream), and stronger power over the left hemisphere when the context comprised large Δf_1 (participants reported hearing two streams). Differences in alpha synchronization have been reported for bistable stimuli,²⁹ auditory illusions,³⁰ and during binocular rivalry,¹⁹ with greater alpha power to fragmented than to integrated visual objects.¹⁸

In the present study, contextual sequences changed the brain state, as indexed by oscillatory activity, such that the subsequent construal of an identical bistable stimulus is opposite to that of the previous stimulus. This contrast effect could be related to neural adaptation in auditory cortices.^{5,11} The neural adaptation account has received some support from our initial ERP analysis, revealing a contrast effect associated with changes in neural activity localized in the auditory cortex.¹⁴ In that study, the activity in the auditory cortex correlated with the changes in Δf_1 . However, the relationship between neural activity and perception of the ambiguous sequence was not examined, making it difficult to determine whether the change in neural activity localized in the auditory cortex was associated with interpreting the bistable stimulus. In the present study, the contrast effect was related to changes in alpha power, with source activity encompassing the superior temporal gyrus. Although the latter appears consistent with the neural adaptation account, only activity from the right sensorimotor areas was correlated with the contrast effect. This suggests that in addition to neural adaptation, other processes, such as attention, may be contributing.

Alpha power has been used in many studies as a neural metric of visual^{19,20,31–33} and auditory attention.^{34–39} Given previous studies associating alpha power with attention, the observed changes in these rhythms could be associated with an attentional bias toward the percept opposite to that of the prior stimulus. One source of alpha activity was localized in posterior parietal regions, consistent with attentional processing and perceptual switching.¹⁹ The contrast effect may be related to perceptual adaptation, with the listener's attention being biased toward the other construal in the context of bistable stimuli. Participants may also compare the incoming stimuli sequence with the sequence's representation, making the incoming stand out in the opposite perceptual space of the previous stimulus.

In the present study, we observe perception-related changes in alpha power, with greater power elicited by the bistable sequence when preceded by the same sequence but with larger Δf_1 . The greater alpha synchrony could indicate that participants' attention is more focused on the galloping rhythm rather than being divided between the two streams of vowels. This would be consistent with prior research showing a difference in alpha power between focused and divided attention.⁴⁰ Findings from our source analysis are also compatible with the role of attention in bistable perception. Prior research using fMRI has shown activity in parietal regions associated with bistable perception during auditory stream segregation.^{5,6} Alpha desynchronization may index reflective attention to the dominant percept (galloping rhythm versus a single stream of vowels), bringing this percept into the foreground.²⁹ In the present study, alpha power may reflect the prioritization of one percept over another,⁴¹ with greater alpha synchrony over the left hemisphere meaning less salient or "silent" representation of two vowel stream percepts.

Source localization of these alpha power effects suggests that the contrast effect of bistable vowel sequences engages left lateralized posterior speech areas and bilateral sensorimotor areas. The angular gyrus and adjacent areas play an important role in speech perception⁴² and audio-vocal sensory-motor transformations^{43,44} and are sensitive to manipulations promoting the speech continuity illusion⁴⁵ and different perceptual interpretations of ambiguous speech sounds with varying stimulus context.⁴⁶ Our findings are also consistent with prior fMRI studies showing activity in inferior parietal lobule and areas near or including the angular gyrus during bistable perception.⁴⁷ The difference in source activity located in the left and right sensorimotor areas may reflect response preparation associated with the dominant percept or

Table 1. Brain-behavior correlations

		Contrast Effect	Left Angular Gyrus	Left Sensorimotor	Right Sensorimotor
Contrast Effect	r	1			
	p value	15			
	N				
Left Angular Gyrus	r	−0.430	1		
	p value	0.128			
	N	15			
Left Sensorimotor	r	−0.411	0.578 ^a	1	
	p value	0.128	0.050		
	N	15	15		
Right Sensorimotor	r	−0.650 ^a	0.659 ^a	0.538	1
	p value	0.024	0.024	0.057	
	N	15	15	15	

^aSignificant correlation ($p < 0.05$, two-tailed corrected for multiple comparisons).

activation in speech motor areas associated with processing the two different vowels.⁴⁸ The significant correlation between the source activity in the right sensorimotor area and the behavioral contrast effect is also consistent with the aforementioned hypotheses that the contrast effect is related to a bias when comparing the incoming stimulus with the previous one. Together, these results suggest that the contrast effect during bistable perception of vowel sequences is associated with auditory-motor integration.

Limitations of the study

While our sample size was sufficient to observe reliable differences in oscillatory activity, the source localization of neural alpha oscillations may change with the number of participants. A larger sample would enhance the source model adequacy and may reveal different and/or additional regions. The present study also focused primarily on one context-driven perception of bistable ambiguous stimuli with two alternatives. It remains to be seen whether the same brain areas would be involved when the perception switches back and forth between alternatives. Consequently, future studies could expand on the findings reported here and compare neural oscillations when the prior context yields the same and opposite percept to that experienced from the previous stimulus. Lastly, the exact extent of the brain areas associated with the contrast effect should be interpreted cautiously because source reconstructions depend on many factors that can influence the number and locations of sources (e.g., head model, regularization, voxel size, etc.). The findings from the present study should be confirmed in future studies using methods with better spatial resolution, such as subdural EEG recording and fMRI.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [RESOURCE AVAILABILITY](#)
 - Lead contact
 - Material availability
 - Data and code availability
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
 - Participants
- [METHOD DETAILS](#)
 - Stimuli and task
 - Recording of neuroelectric brain activity
 - Time-frequency analysis
 - Multiple source beamformer (MSFB)
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)

ACKNOWLEDGMENTS

This research was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant to Claude Alain [RGPIN- 2021–02721].

AUTHOR CONTRIBUTIONS

Conceptualization, C.A., G.M.B., and J.S.S.; methodology, C.A., G.M.B., and J.S.S.; investigation, C.A., D.S., K.G., and L.J.B; writing – original draft, C.A. and L.J.B; writing – review & editing, C.A., L.J.B., G.M.B., D.S., and J.S.S; funding acquisition, C.A.; resources, C.A., M.G.B., and J.S.S.; supervision, C.A., M.G.B., and J.S.S.

DECLARATION OF INTERESTS

Competing financial interests: The authors declare no competing financial interests.

Received: May 11, 2023

Revised: October 5, 2023

Accepted: November 11, 2023

Published: November 14, 2023

REFERENCES

- Denham, S.L., and Winkler, I. (2006). The role of predictive models in the formation of auditory streams. *J. Physiol. Paris* 100, 154–170.
- Snyder, J.S., and Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychol. Bull.* 133, 780–799.
- Schadwinkel, S., and Gutschalk, A. (2011). Transient bold activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus. *J. Neurophysiol.* 105, 1977–1983.
- Kondo, H.M., and Kashino, M. (2009). Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *J. Neurosci.* 29, 12695–12701.
- Higgins, N.C., Scurry, A.N., Jiang, F., Little, D.F., Alain, C., Elhilali, M., and Snyder, J.S. (2023). Adaptation in the sensory cortex drives bistable switching during auditory stream segregation. *Neurosci. Conscious.* 2023, niac019.
- Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *J. Cogn. Neurosci.* 17, 641–651.
- Hill, K.T., Bishop, C.W., Yadav, D., and Miller, L.M. (2011). Pattern of BOLD signal in auditory cortex relates acoustic response to perceptual streaming. *BMC Neurosci.* 12, 85.
- Bernstein, L.J., and Cooper, L.A. (1997). Direction of motion influences perceptual identification of ambiguous figures. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 721–737.
- Hochberg, J., and Peterson, M.A. (1987). Piecemeal organization and cognitive components in object perception: perceptually coupled responses to moving objects. *J. Exp. Psychol. Gen.* 116, 370–380.
- Goldstein, A.G. (1958). On the after-effects of the waterfall and spiral illusions. *Am. J. Psychol.* 71, 608–609.
- Snyder, J.S., Carter, O.L., Hannon, E.E., and Alain, C. (2009). Adaptation reveals multiple levels of representation in auditory stream segregation. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1232–1244.
- Snyder, J.S., Carter, O.L., Lee, S.K., Hannon, E.E., and Alain, C. (2008). Effects of context on auditory stream segregation. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1007–1016.
- Snyder, J.S., Holder, W.T., Weintraub, D.M., Carter, O.L., and Alain, C. (2009). Effects of prior stimulus and prior perception on neural correlates of auditory stream segregation. *Psychophysiology* 46, 1208–1215.
- Alain, C., Arseneault, J.S., Garami, L., Bidelman, G.M., and Snyder, J.S. (2017). Neural Correlates of Speech Segregation Based on Formant Frequencies of Adjacent Vowels. *Sci. Rep.* 7, 40790.
- Huguet, G., Rinzel, J., and Hupé, J.M. (2014). Noise and adaptation in multistable perception: noise drives when to switch, adaptation determines percept choice. *J. Vis.* 14, 19.
- Little, D.F., Snyder, J.S., and Elhilali, M. (2020). Ensemble modeling of auditory streaming reveals potential sources of bistability across the perceptual hierarchy. *PLoS Comput. Biol.* 16, e1007746.
- Rankin, J., Sussman, E., and Rinzel, J. (2015). Neuromechanistic Model of Auditory Bistability. *PLoS Comput. Biol.* 11, e1004555.
- Flevaris, A.V., Martínez, A., and Hillyard, S.A. (2013). Neural substrates of perceptual integration during bistable object perception. *J. Vis.* 13, 17.
- Drew, A., Torralba, M., Ruzzoli, M., Moris Fernández, L., Sabaté, A., Pápai, M.S., and Soto-Faraco, S. (2022). Conflict monitoring and attentional adjustment during binocular rivalry. *Eur. J. Neurosci.* 55, 138–153.
- Piantoni, G., Romeijn, N., Gomez-Herrero, G., Van Der Werf, Y.D., and Van Someren, E.J.W. (2017). Alpha Power Predicts Persistence of Bistable Perception. *Sci. Rep.* 7, 5208.
- Katyal, S., He, S., He, B., and Engel, S.A. (2019). Frequency of alpha oscillation predicts individual differences in perceptual stability during binocular rivalry. *Hum. Brain Mapp.* 40, 2422–2433.
- Kornmeier, J., and Bach, M. (2012). Ambiguous figures - what happens in the brain when perception changes but not the stimulus. *Front. Neurosci.* 6, 51.
- Cao, Y., Oostenveld, R., Alday, P.M., and Piai, V. (2022). Are alpha and beta oscillations spatially dissociated over the cortex in context-driven spoken-word production? *Psychophysiology* 59, e13999.
- Crespo-García, M., Pinal, D., Cantero, J.L., Díaz, F., Zurrón, M., and Atienza, M. (2013). Working memory processes are mediated by local and long-range synchronization of alpha oscillations. *J. Cogn. Neurosci.* 25, 1343–1357.
- Tzagarakis, C., West, S., and Pellizzer, G. (2015). Brain oscillatory activity during motor preparation: effect of directional uncertainty on beta, but not alpha, frequency band. *Front. Neurosci.* 9, 246.
- Bregman, A.S. (1978). Auditory streaming is cumulative. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 380–387.
- Deike, S., Heil, P., Böckmann-Barthel, M., and Brechmann, A. (2012). The Build-up of Auditory Stream Segregation: A Different Perspective. *Front. Psychol.* 3, 461.
- Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., and Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proc. Natl. Acad. Sci. USA* 98, 694–699.
- Torralba Cuello, M., Drew, A., Sabaté San José, A., Moris Fernández, L., and Soto-Faraco, S. (2022). Alpha fluctuations regulate the accrual of visual information to awareness. *Cortex* 147, 58–71.
- Shahin, A.J. (2019). Neural evidence accounting for interindividual variability of the McGurk illusion. *Neurosci. Lett.* 707, 134322.
- Foxe, J.J., and Snyder, A.C. (2011). The Role of Alpha-Band Brain Oscillations as a Sensory Suppression Mechanism during Selective Attention. *Front. Psychol.* 2, 154.
- Snyder, A.C., and Foxe, J.J. (2010). Anticipatory attentional suppression of visual features indexed by oscillatory alpha-band power increases: a high-density electrical mapping study. *J. Neurosci.* 30, 4024–4032.
- Van Diepen, R.M., Foxe, J.J., and Mazaheri, A. (2019). The functional role of alpha-band activity in attentional processing: the current zeitgeist and future outlook. *Curr. Opin. Psychol.* 29, 229–238.
- Ahveninen, J., Seidman, L.J., Chang, W.T., Hämäläinen, M., and Huang, S. (2017). Suppression of irrelevant sounds during auditory working memory. *Neuroimage* 161, 1–8.
- Jäncke, L., Leipold, S., and Burkhard, A. (2018). The neural underpinnings of music listening under different attention conditions. *Neuroreport* 29, 594–604.
- Shen, D., Ross, B., and Alain, C. (2016). Temporal cuing modulates alpha oscillations during auditory attentional blink. *Eur. J. Neurosci.* 44, 1833–1845.
- Teoh, E.S., and Lalor, E.C. (2019). EEG decoding of the target speaker in a cocktail party scenario: considerations regarding dynamic switching of talker location. *J. Neural. Eng.* 16, 036017.
- Wöstmann, M., Alavash, M., and Obleser, J. (2019). Alpha Oscillations in the Human Brain Implement Distractor Suppression

- Independent of Target Selection. *J. Neurosci.* 39, 9797–9805.
39. Backer, K.C., Binns, M.A., and Alain, C. (2015). Neural dynamics underlying attentional orienting to auditory representations in short-term memory. *J. Neurosci.* 35, 1307–1318.
 40. Sokoliuk, R., Mayhew, S.D., Aquino, K.M., Wilson, R., Brookes, M.J., Francis, S.T., Hanslmayr, S., and Mullinger, K.J. (2019). Two Spatially Distinct Posterior Alpha Sources Fulfill Different Functional Roles in Attention. *J. Neurosci.* 39, 7183–7194.
 41. Weisz, N., Kraft, N.G., and Demarchi, G. (2020). Auditory cortical alpha/beta desynchronization prioritizes the representation of memory items during a retention period. *Elife* 9, e55508.
 42. Dai, G., Chen, M., Chen, X., Guo, Z., Li, T., Jones, J.A., Wu, X., Li, J., Liu, P., Liu, H., and Liu, D. (2022). A causal link between left supplementary motor area and auditory-motor control of vocal production: Evidence by continuous theta burst stimulation. *Neuroimage* 264, 119767.
 43. Pa, J., and Hickok, G. (2008). A parietal-temporal sensory-motor integration area for the human vocal tract: evidence from an fMRI study of skilled musicians. *Neuropsychologia* 46, 362–368.
 44. Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
 45. Shahin, A.J., Bishop, C.W., and Miller, L.M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *Neuroimage* 44, 1133–1143.
 46. Bidelman, G.M., Pearson, C., and Harrison, A. (2021). Lexical Influences on Categorical Speech Perception Are Driven by a Temporoparietal Circuit. *J. Cogn. Neurosci.* 33, 840–852.
 47. Bisenius, S., Trapp, S., Neumann, J., and Schroeter, M.L. (2015). Identifying neural correlates of visual consciousness with ALE meta-analyses. *Neuroimage* 122, 177–187.
 48. Jenson, D., Harkrider, A.W., Thornton, D., Bowers, A.L., and Saltuklaroglu, T. (2015). Auditory cortical deactivation during speech production and following speech perception: an EEG investigation of the temporal dynamics of the auditory alpha rhythm. *Front. Hum. Neurosci.* 9, 534.
 49. Hutka, S.A., Alain, C., Binns, M.A., and Bidelman, G.M. (2013). Age-related differences in the sequential organization of speech sounds. *J. Acoust. Soc. Am.* 133, 4177–4187.
 50. Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190.
 51. Maris, E., Schoffelen, J.M., and Fries, P. (2007). Nonparametric statistical testing of coherence differences. *J. Neurosci. Methods* 163, 161–175.
 52. Hochberg, Y., and Benjamini, Y. (1990). More powerful procedures for multiple significance testing. *Stat. Med.* 9, 811–818.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Original data and code for analyses	https://doi.org/10.5683/SP3/3XM0XY	na
Software and algorithms		
BESA Research 7.1	https://www.besa.de/	RRID:SCR_009530
BESA Statistic 2.1	https://www.besa.de/	RRID:SCR_009530
EEGLab	https://sccn.ucsd.edu/eeglab/index.php	RRID:SCR_007292
IBM SPSS 28	https://www.ibm.com/products/spss-statistics	RRID:SCR_002865
MATLAB Version 2021	https://it.mathworks.com/products/matlab.html	RRID:SCR_001622

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Claude Alain (calain@research.baycrest.org).

Material availability

This study did not generate new unique reagents.

Data and code availability

The data that support the findings of this study are available from the [lead contact](#) upon reasonable request. All original data and codes have been deposited at Borealis: The Canadian Dataverse Repository and are publicly available as of the publication date. DOI is listed in the [key resources table](#). Statistical analyses were run using BESA Statistic and SPSS. The [lead contact](#) will provide any additional information needed to reanalyze the data reported in the paper.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants

Sixteen participants were included in our original study.¹⁴ In contrast to that study, the current study uses a much longer epoch in order to examine bistable perception (−30 to 500 ms versus −1000 to 7500 ms, respectively). Due to the length of the longer epoch (see below), electrophysiological data from one participant had to be excluded because of the high number of vertical eye movements and blinks contaminating the responses of interest. EEG data from 15 participants were analyzed ($M_{\text{age}} = 23.53$ years, $SD = 4.33$; seven females). All but one participant was right-handed. All were fluent English speakers with no known neurological, psychiatric, hearing or speech disorders. The study was carried out in accordance with relevant guidelines and regulations and was approved by the University of Toronto and Baycrest Hospital Human Subject Review and Research Ethics Committees. Participants gave informed written consent and received a small honorarium for participation.

METHOD DETAILS

Stimuli and task

Vowel sounds /i/ (as in “see”) and /ae/ (as in “cat”), hereafter referred to as “ee” and “ae” were synthesized using a cascade formant synthesizer implemented in MATLAB using a sampling rate of 48828 Hz. Both vowels were 100 ms long and had a voice fundamental frequency (f_0) of 100 Hz. The first formant frequency (f_1) of the vowel /ee/ was fixed at 400 Hz, whereas the f_1 of the vowel /ae/ was allowed to vary. The difference in f_1 (Δf_1) between /ee/ and /ae/ was small ($\Delta f_1 = 47$ Hz), intermediate ($\Delta f_1 = 110$ Hz), or large ($\Delta f_1 = 285$ Hz). A prior study using these values shows that participants consistently report hearing one stream, report either one or two streams (ambiguous sequence) and generally report two streams, respectively.⁴⁹ Stimuli were presented binaurally at 75 decibels (dB) sound pressure level through Sennheiser HD 265 headphones.

Vowels were presented in an ABA- pattern as /ee/-/ae/-/ee/ (Figure 1A). Only the second vowel within the triplet was manipulated. The inter-stimulus interval (ISI) between /ee/ and /ae/ was a fixed 50 ms, and the ISI between triplets was 100 ms. Each trial consisted of a

contextual cue sequence followed by a sequence with an intermediate Δf_1 , referred to as the ambiguous sequence (Figure 1B). For the contextual cue sequence, the Δf_1 was small, intermediate, or large. In the ambiguous test sequence, the Δf_1 was always intermediate. The contextual cue and the ambiguous sequences contained 14 repetitions of the /ee/-/ae/-/ee/- triplets and lasted 6.90 seconds (sec). They were separated by 1.44 sec of silence. After the contextual cue and ambiguous sequences, participants pressed one of two keys on a response box (Tucker-Davis Technologies) indicating whether they heard one galloping rhythm or two concurrent streams. Participants had 3.50 sec to respond, and then the next trial began. Participants completed five blocks of 30 trials each for a total of 150 trials, with each Δf_1 condition (small, intermediate, large) being presented 50 times. In each block of trials, the three levels of Δf_1 were presented in random order. The experiment took about one hour to complete including breaks between blocks of trials.

Recording of neuroelectric brain activity

The EEG was digitized continuously (sampling rate 500 Hz; online bandpass filter of 0.05–100 Hz) from an array of 64 electrodes using NeuroScan Synamps2 (Compumedics, El Paso, TX, USA). Eye movements were monitored with electrodes at the outer canthi and below the eyes. All electrodes were referenced to the vertex electrode (i.e., Cz) during recording. For offline data analysis, the EEG signal at each electrode was re-referenced to the average signal of all electrodes (i.e., average reference).

For each participant, a set of ocular movements was identified from the continuous EEG recording and then used to generate spatial components that best account for blinks and vertical and lateral eye movements using BESA Research software (version 7.1). The spatial topographies were then subtracted from the continuous EEG to correct for eye movements and eye blinks. After correcting for eye movements, recordings were automatically scanned for artifacts. Epochs (–1 sec to 16 sec) including deflections exceeding peak-to-peak 120 μ V were marked and excluded from the analysis. The number of trials included in the averages ranged from 38 to 50, i.e., at least 75% of trials for each condition without deflection exceeding 120 μ V.

Time-frequency analysis

The time-frequency analysis of the EEG signal power was performed with the BESA Research software (version 7.1). The continuous EEG was segmented in epochs consisting of 1 sec of pre-stimulus activity and 16 sec of post-stimulus activity time-locked to the onset of the sequence. The baseline duration was chosen to provide a good estimate of non-task specific oscillatory power while minimizing the impact of response-related activity from the previous trial. A complex demodulation method with 1 Hz wide frequency bins and 50 ms time resolution was used to decompose the single-trial EEG data into frequency representation from 2 to 50 Hz.

Multiple source beamformer (MSFB)

A time-frequency source localization approach (multiple source beamformer in BESA version 7.1)²⁸ was applied to a 1000ms time window centered on the peak of the difference in alpha power (i.e., 12.6 to 13.6 sec) from the time-frequency analysis. This technique spatially filters scalp-recorded EEG data to estimate the source power of specific location(s) in the brain. Source power distributions in the 3D brain are estimated by iteratively constructing the beamformer at each voxel. We used a standardized FEM (finite element method) model created from an average head using 50 individual MRIs in Talairach space provided by BESA Research 7.1, with a voxel size in Talairach space of 9 mm and a regularization of 0.001.

QUANTIFICATION AND STATISTICAL ANALYSIS

The behavioral outcome measure was the proportion of trials in which participants reported hearing two streams after the contextual cue and the ambiguous sequences. The effect of contextual cue (i.e., Δf_1) on the perception of the ambiguous sequence was quantified by comparing the proportion of trials in which participants reported hearing two auditory streams using a repeated measures analysis of variance with Δf_1 as the within-subject factor.

The effect of the contextual cue on oscillatory activity was examined using clustered-based statistics and permutation tests (BESA Statistics 2.1). First, t-tests were used to identify clusters in time (adjacent time points) and space (adjacent electrodes) where oscillatory activity differed between conditions. The channel diameter was 4 cm, allowing up to four electrode neighbours per analysis cluster. We used an alpha level of 0.05 for cluster building to analyze the time-frequency data. We used an alpha level of 0.01 for cluster building to analyze oscillatory source activity. This more stringent alpha level for cluster building allows for separation from nearby sources. A Monte-Carlo resampling technique^{50,51} was then used to identify the clusters with higher values than 95% (one-sided t-test) of all clusters derived by random permutation of the data. The number of permutations was set at 5,000. Importantly, this procedure corrects for multiple comparisons over time and electrodes to minimize false positives.^{50,51}

Brain-behaviour relationships were examined using Pearson correlations. We used the Benjamini-Hochberg method to adjust the family-wise p-value for multiple comparisons with $q = 0.1$, $m = 6$ (i.e., the total number of p values) and $p = 0.05$.⁵²