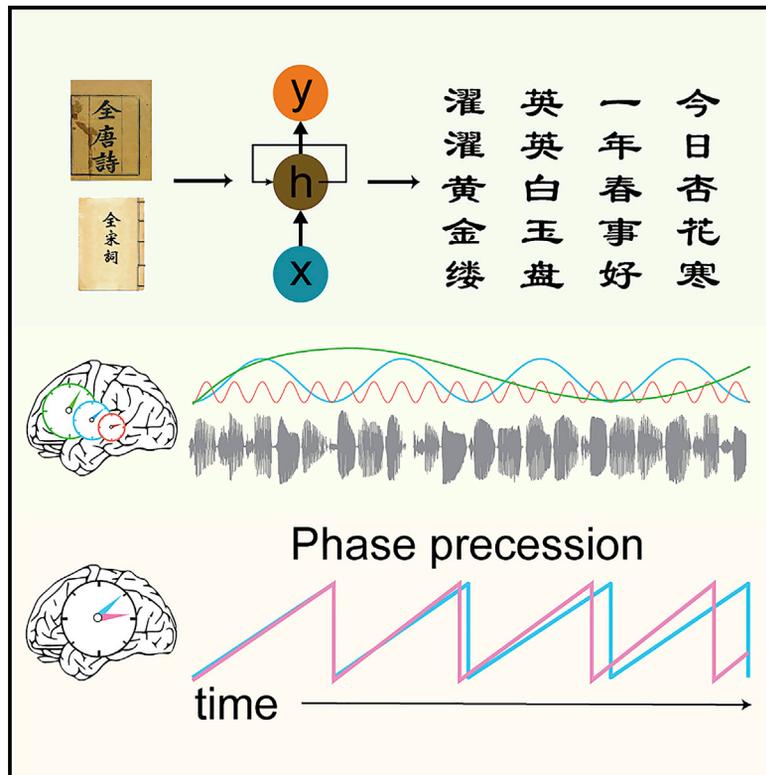


# Current Biology

## Constrained Structure of Ancient Chinese Poetry Facilitates Speech Content Grouping

### Graphical Abstract



### Authors

Xiangbin Teng, Min Ma, Jinbiao Yang, Stefan Blohm, Qing Cai, Xing Tian

### Correspondence

xing.tian@nyu.edu

### In Brief

Teng et al. combine recurrent neural network and neurophysiology to investigate how the structured format of poetry aids speech perception. The structured poetic format renders poems predictable across multiple timescales and facilitates speech segmentation. Phase precession indicating predictive processes in speech segmentation is observed.

### Highlights

- Listeners parse ancient Chinese poetry according to the poetic structures
- Two neural timescales reflect the grouping of distinct poetic features
- Phase precession is observed in speech segmentation, indicating predictive processes
- Co-occurrence of words does not necessarily facilitate speech segmentation



# Constrained Structure of Ancient Chinese Poetry Facilitates Speech Content Grouping

Xiangbin Teng,<sup>1,8</sup> Min Ma,<sup>2,8</sup> Jinbiao Yang,<sup>3,4,5,6</sup> Stefan Blohm,<sup>1</sup> Qing Cai,<sup>4,7</sup> and Xing Tian<sup>3,4,7,9,\*</sup>

<sup>1</sup>Department of Neuroscience, Max Planck Institute for Empirical Aesthetics, Frankfurt 60322, Germany

<sup>2</sup>Google Inc., 111 8th Avenue, New York, NY 10010, United States

<sup>3</sup>Division of Arts and Sciences, New York University Shanghai, Shanghai 200122, China

<sup>4</sup>NYU-ECNU Institute of Brain and Cognitive Science at NYU Shanghai, Shanghai 200062, China

<sup>5</sup>Max Planck Institute for Psycholinguistics, Wundtlaan 1, Nijmegen 6525 XD, the Netherlands

<sup>6</sup>Centre for Language Studies, Radboud University, Erasmusplein 1, Nijmegen 6525 HT, the Netherlands

<sup>7</sup>Key Laboratory of Brain Functional Genomics (MOE & STCSM), Shanghai Changning-ECNU Mental Health Center, Institute of Cognitive Neuroscience, School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China

<sup>8</sup>These authors contributed equally

<sup>9</sup>Lead Contact

\*Correspondence: [xing.tian@nyu.edu](mailto:xing.tian@nyu.edu)

<https://doi.org/10.1016/j.cub.2020.01.059>

## SUMMARY

Ancient Chinese poetry is constituted by structured language that deviates from ordinary language usage [1, 2]; its poetic genres impose unique combinatory constraints on linguistic elements [3]. How does the constrained poetic structure facilitate speech segmentation when common linguistic [4–8] and statistical cues [5, 9] are unreliable to listeners in poems? We generated artificial *Jueju*, which arguably has the most constrained structure in ancient Chinese poetry, and presented each poem twice as an isochronous sequence of syllables to native Mandarin speakers while conducting magnetoencephalography (MEG) recording. We found that listeners deployed their prior knowledge of *Jueju* to build the line structure and to establish the conceptual flow of *Jueju*. Unprecedentedly, we found a phase precession phenomenon indicating predictive processes of speech segmentation—the neural phase advanced faster after listeners acquired knowledge of incoming speech. The statistical co-occurrence of monosyllabic words in *Jueju* negatively correlated with speech segmentation, which provides an alternative perspective on how statistical cues facilitate speech segmentation. Our findings suggest that constrained poetic structures serve as a temporal map for listeners to group speech contents and to predict incoming speech signals. Listeners can parse speech streams by using not only grammatical and statistical cues but also their prior knowledge of the form of language.

## RESULTS AND DISCUSSION

### Jueju and Poem Generation

An example of *Jueju* is shown in Figure 1A. We generated 150 artificial poems that resemble *Jueju* using a recurrent neural

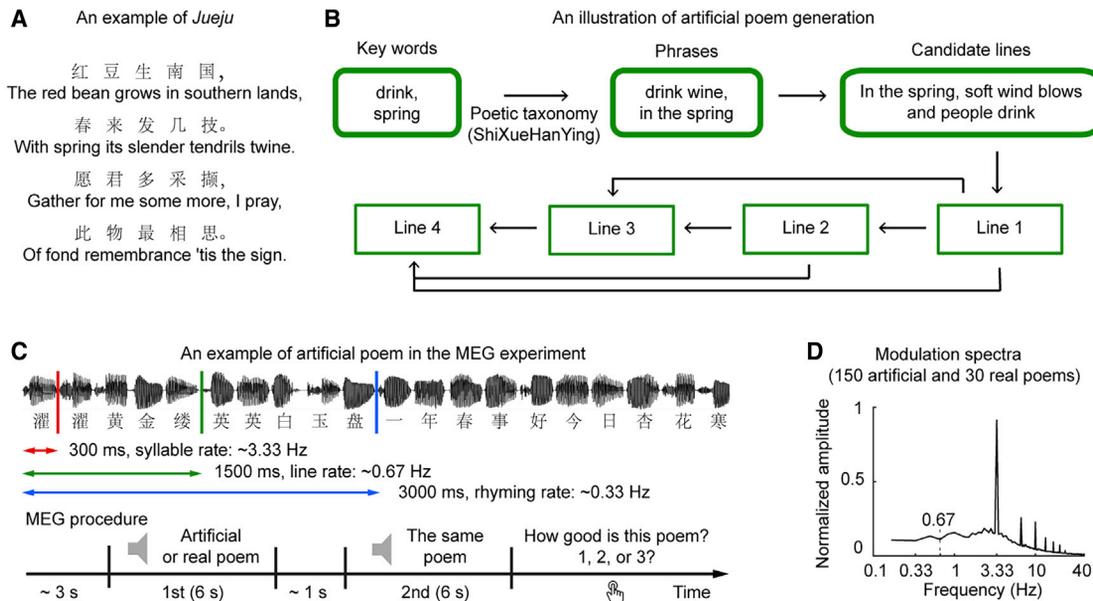
network model trained on a large corpus of ancient Chinese poems (see Figure 1B and STAR Methods), which allowed us to control semantic and thematic content, linguistic choices (i.e., style), and listeners' familiarity. We further selected 30 real poems from the poem corpus, which matched the generated poems in regard with listeners' familiarity and linguistic choices. Phrases in one line of *Jueju* are not strictly constrained by grammatical rules; lines are paired without strict semantic coherence [1, 2]. It is challenging for Chinese listeners to construct structures of unfamiliar *Jueju* simply based on semantic scheme and grammatical rules.

To eliminate the confounding acoustic cues (e.g., pauses and intonation), we synthesized each monosyllabic word individually and concatenated them to construct the auditory stimuli as isochronous sequences of syllables (Figure 1C). Each syllable was 300 ms (syllable rate:  $\sim 3.33$  Hz); each line of a poem contained five syllables, resulting in a duration of 1,500 ms (line rate:  $\sim 0.67$  Hz). A couplet of lines contained two lines (3,000 ms) and closed with a rhyme word (rhyming rate:  $\sim 0.33$  Hz). The modulation spectra of the stimuli (Figure 1D) demonstrated no salient amplitude modulations around the line rate or the rhyming rate.

### Speech Segmentation Driven by Listeners' Prior Knowledge of Poetic Structure

Thirteen native speakers of Mandarin Chinese, who received formal education in China and learned the structure of *Jueju*, listened to each poem twice while undergoing magnetoencephalography (MEG) recording. We hypothesized that the neural signatures at the timescales corresponding to the line and rhyming rates would be observed when listeners were actively parsing the poetic speech stream according to their prior knowledge of the structure of *Jueju*. We first conducted a spectral analysis on the averaged MEG responses across two presentations in the independently defined auditory channels (Figure 2A and STAR Methods). The amplitude spectrum shows prominent peaks corresponding to the syllable rate and its harmonics, which reflects neural tracking of the amplitude envelopes of the stimuli (Figure 1D). After removing the intrinsic 1/f amplitude





**Figure 1. Poem Generation and Experimental Paradigm**

(A) An example of *Jueju*. *Jueju* contains four lines with five characters in each line. English sense-for-sense translations are shown under each line [10].  
 (B) Poem generation using recurrent neural network. To generate the first line, keywords were chosen to generate phrases by referring to Chinese poetic taxonomy (*ShiXueHanYing*). Phrases were combined into lines, from which the first line of a poem was selected. Subsequent lines were generated based on all the previously generated lines, which are subject to phonological (rhyme) and structural constraints (the length of each line).  
 (C) MEG experimental paradigm. Each poem stimulus was presented as an isochronous sequence of auditory syllables twice during MEG recording. At the top, an example artificial poem and its synthesized waveform are shown. Each syllable is 300 ms long (syllable rate:  $\sim 3.33$  Hz), each line is 1,500 ms long (line rate:  $\sim 0.67$  Hz), and the rhyming rate between two lines is  $\sim 0.33$  Hz. The participants were required to evaluate the poem by the end of each trial.  
 (D) Modulation spectra of the poem stimuli. The modulation spectra of 150 artificial poem stimuli and 30 real poem stimuli show salient peaks at the syllable rate and its harmonics, but not at the line rate or the rhyming rate. See also Figure S1.

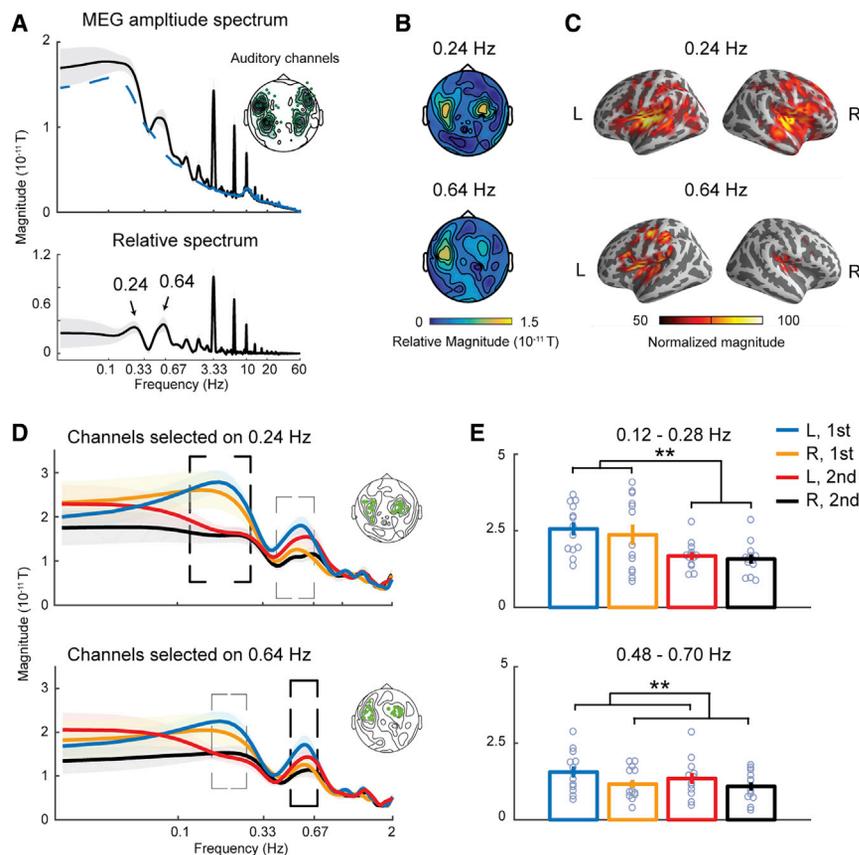
distribution estimated using a surrogate permutation procedure, two spectral components emerged in the relative amplitude spectrum (Figure 2A, lower panel), with  $\sim 0.64$  Hz matching the line rate. The topographies and source localizations of the two components are shown in Figures 2B and 2C.

We next examined the effects of hemispheric lateralization and repetition of the poem stimuli using ten MEG channels for each hemisphere and each frequency component, which were selected based on the topographies of the spectral components of  $\sim 0.24$  and  $\sim 0.64$  Hz, respectively (Figure 2B). We first conducted a cluster-based test to determine frequency ranges showing significant effects ( $p < 0.01$ ) (Figure 2D). Through a two-way repeated-measure ANOVA (rmANOVA) within each significant frequency range (Figure 2E), we found that the amplitude of the second presentation of the poem stimuli significantly decreased compared with the first presentation between 0.12 and 0.28 Hz ( $F(1,12) = 9.45$ ,  $p = 0.0095$ ,  $\eta_p^2 = 0.441$ )—a phenomenon of repetition suppression on the level of the entire poem stimuli [11], which was arguably induced by the expectation of incoming stimuli [12]. The main effect of lateralization and the interaction effect were not significant ( $F(1,12) = 0.60$ ,  $p = 0.452$ ,  $\eta_p^2 = 0.048$ ;  $F(1,12) = 0.10$ ,  $p = 0.756$ ,  $\eta_p^2 = 0.008$ ). The significant left lateralization was observed between 0.48 and 0.70 Hz ( $F(1,12) = 9.58$ ,  $p = 0.009$ ,  $\eta_p^2 = 0.444$ ), which is consistent with previous literature showing a left lateralization of sentence-level processing [13–17]. The main effect of repetition and the interaction effect were not significant ( $F(1,12) =$

1.50,  $p = 0.244$ ,  $\eta_p^2 = 0.111$ ;  $F(1,12) = 1.56$ ,  $p = 0.236$ ,  $\eta_p^2 = 0.115$ ). Interestingly, this finding suggests that the left-lateralized sentence-level process does not necessarily involve strict syntactical processing but represents more general combinatory processes of linguistic elements, as the lines of *Jueju* lack well-constructed sentence structures. These findings cannot be explained by the acoustic properties of the poem stimuli (Figure S1). Moreover, the number of channels included in the analyses did not affect our observations (Figures S2A and S2B).

These results demonstrate that listeners employ the poetic structure as a template to actively group words into cohorts congruent to the formal and conceptual structure of *Jueju*, which is complementary to previous findings that neural signals mark boundaries of linguistic elements, such as words and phrases [4–8]. This structure-based combinatory process potentially explains how the constrained structure aids poem appreciation—as poetic languages deviate from ordinary languages and unique combinations of words are often used, imposing a stringent formal and conceptual structure provides a poetic temporal frame for listeners to constitute semantic cohorts intended by poets.

We did not find any neural evidence of the neural tracking of the rhyming structure, although the neural network took into account the phonological rules [18]. This could be because coarticulations between words and prosodic cues produced by human speakers important to rhyme were absent in the synthesized poem stimuli. Another possibility is that the rhyme relation



In the frequency range between 0.48 and 0.70 Hz (lower panel), significant left lateralization was found ( $p < 0.01$ ). The shaded areas in (A) and (D) represent one standard error of the mean; the error bars in (E) represent one standard error of the mean. See also [Figures S1](#) and [S2](#).

can only be identified after the occurrence of the last word, which may not promote the grouping within a couplet.

### Neural Phase Precession Induced by the Repetition of Poem Stimuli

Distinct neural signatures are observed previously at boundaries between complex linguistic structures in the temporal domain [5, 7, 19]; the temporal information of the neural signals may reveal the active operation of the poem parsing due to listeners' prior knowledge of poems. We filtered the MEG signals between 0.1 and 2 Hz to remove components ( $\sim 3.33$  Hz) evoked by the amplitude envelopes. The filtered temporal responses showed robust onset and offset responses to the poem stimuli ([Figure 3A](#)). Clear peaks were observed at the boundary between the first and second lines ( $\sim 1.5$  s), as well as between the second and third lines ( $\sim 3.0$  s); the peak around the boundary between the third and the fourth lines seemingly shifted forward. The fact that the time points of peaks closely matched the line boundaries suggests that the brain may actively predict the line structure to segment the poem contents.

We selected two frequency bands centered around  $\sim 0.24$  and  $\sim 0.64$  Hz, separately ([Figure 3B](#), upper panel; [STAR Methods](#)). The neural dynamics between 0.1 and 0.4 Hz primarily reflected the onset and offset responses to the poem stimuli as well as the neural responses to the boundary between the third and fourth lines around 4.5 s. The temporal dynamics in the

frequency band between 0.4 and 0.96 Hz matched the line rate and a peak-to-peak cycle roughly corresponding to the duration of a line (1.5 s). To validate our observation quantitatively and to further examine the origin of the repetition effect at  $\sim 0.24$  Hz ([Figure 2E](#), upper panel), we conducted a series of simulation analyses ([STAR Methods](#)), which demonstrated that the spectral component of  $\sim 0.24$  Hz represented a combination of the onset responses to the poem stimuli and a downward amplitude modulation component ranging from the first line to the third line ([Figure S2C](#)). The repetition effect at  $\sim 0.24$  Hz was associated with the decreased modulation strength in the second presentation of the poem stimuli ([Figure S2D](#)). The downward amplitude modulation across the first three lines indicates a grouping between the first three lines and the last line and is consistent with the conceptual flow of *Jueju*—"Qi (initialization) Cheng (elaboration) Zhuan (continuum/twist) He (integration)," a conceptual organization of the first three lines versus the last line in *Jueju* [20].

We next compared the phase difference between the first and second presentations for each hemisphere in the two frequency bands, respectively. The phase series from the group-averaged responses in [Figure 3B](#) (lower panel) reveals salient phase differences between the two presentations of the same poem stimuli. We then conducted circular statistical tests and found that, in the lower frequency range (0.1–0.4 Hz), the phase of the first presentation led the second presentation in the beginning ( $p < 0.01$ ) ([Figure 3C](#), left panel; left hemisphere:  $-0.6$ – $1.3$  s; right

### Figure 2. Spectral Analyses of MEG Signals

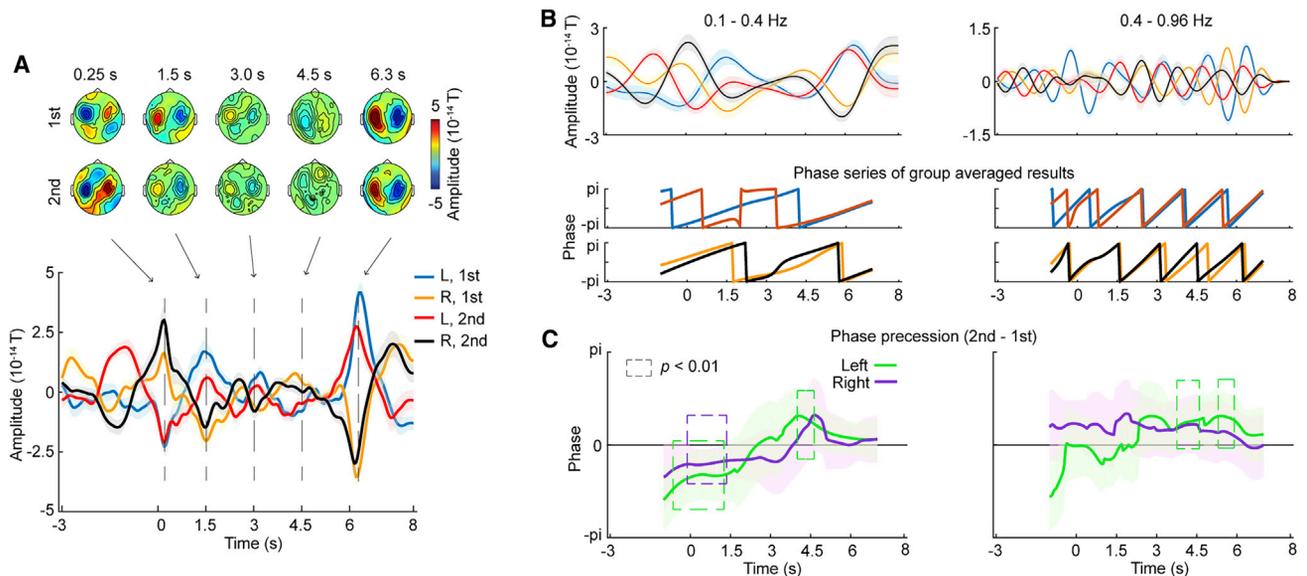
(A) The amplitude spectra of neural responses averaged over the auditory channels (dark green dots on the inserted topographies). The top panel shows the averaged spectrum of both the first and second presentations over 150 poem stimuli. The dashed blue line represents the threshold of the surrogate test (99th percentile). After subtracting the mean of the null distribution out of the MEG spectrum, two spectral components emerged around 0.24 and 0.64 Hz (lower panel), with  $\sim 0.64$  Hz matching the line rate.

(B) MEG topographies of the relative amplitude at 0.24 and 0.64 Hz.

(C) MEG source localization of the spectral components around 0.24 and 0.64 Hz.

(D) Analysis on the channels of interest (ten channels in each hemisphere, green dots on the inserted topographies). The amplitude spectra below 2 Hz are shown for the channels selected based on the components of 0.24 (upper panel) and 0.64 Hz (lower panel), separately. The line color codes for hemispheres and presentations. The dashed boxes indicate the frequency ranges showing significant effects across hemispheres and presentations ( $p < 0.01$ ). The thick dashed box indicates the significant frequency range corresponding to the frequency at which the channels were selected.

(E) Averaged amplitude in the significant frequency ranges. In the low-frequency range between 0.12 and 0.28 Hz (upper panel), the first presentation of the poem stimuli evoked significantly larger amplitude than the second presentation ( $p < 0.01$ ).



**Figure 3. Neural Phase Precession Caused by Poem Repetition**

(A) Neural dynamics of poem parsing. The upper panel shows the topographies of the neural responses to the poem stimuli and the lower panel shows the neural responses between 0.1 and 2 Hz from the channels selected from both the  $\sim 0.24$  and  $\sim 0.64$  Hz components.

(B) The upper panel shows the neural dynamics between 0.1 to 0.4 Hz and between 0.4 to 0.96 Hz; the lower panel shows phase series of averaged neural responses to the poem stimuli at the two frequency ranges.

(C) Phase precession—the phase differences between the first and second presentations in each hemisphere. The colored dashed boxes indicate the periods where the phase difference between the second and first presentations significantly differs from zero in each hemisphere ( $p < 0.01$ ). From 0.1 to 0.4 Hz, the phase series of the second presentation significantly lagged at the beginning in both hemispheres, but flipped to advance faster than that of the first presentation around the onset of the last line in the left hemisphere. From 0.4 to 0.96 Hz, the phase series of the second presentation advanced faster and significantly differed from that of the first presentation after the onset of the third line in the left hemisphere. The shaded areas in (A) and (B) represent one standard error of the mean; the shaded areas in (C) represent one circular standard deviation over participants. See also [Figures S2](#) and [S3](#).

hemisphere:  $-0.05$ – $1.4$  s). An explanation from the perspective of neural signals is simply that the offset responses of the first presentation lasted over 2 s after the offset ([Figure 3A](#)) and affected the onset responses of the second presentations, which caused the phase differences. From a cognitive perspective, the prolonged process of the first presentation likely slowed down the processing of speech contents in the beginning of the second presentation and hence caused the phase delay around the onset. However, the direction of phase difference flipped, and the difference peaked around the transition between the third and last lines (left hemisphere:  $4.05$ – $4.7$  s,  $p < 0.01$ ). The phase series of two presentations aligned with each other as the poem stimuli were reaching the end, which suggests that the brain tracks the unfolding of the poetic stimuli globally.

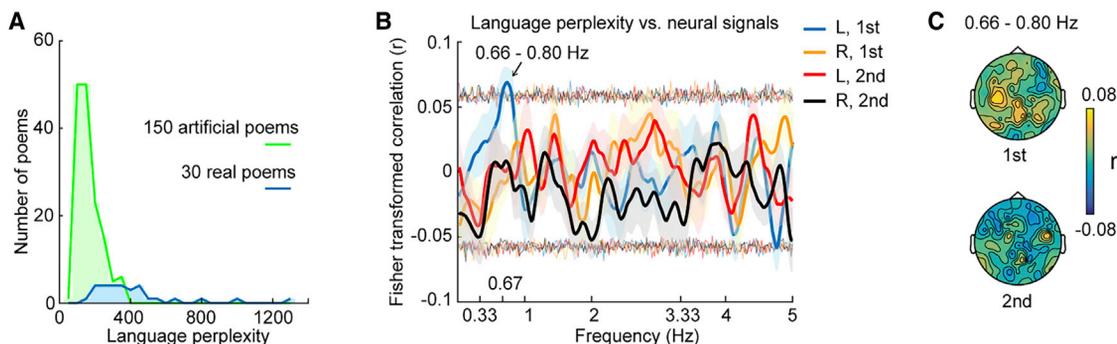
In contrast, the phase series of both presentations were aligned at the beginning in the higher frequency range (0.4–0.96 Hz) ([Figure 3C](#), right panel), but the phase series of the second presentation advanced faster than that of the first presentation after the second line in the left hemisphere ( $3.65$ – $4.8$  s and  $5.05$ – $6.25$  s,  $p < 0.01$ )—a phenomenon of phase precession found in spatial domain [[21](#)], reflecting the prediction of future events [[22](#), [23](#)]. This is probably because, after being primed on the contents of the poem stimuli in the first presentation, the participants registered the words of the first presentation in the working memory and employed the memorized words as temporal landmarks to predict the boundaries of lines in the second presentation, as suggested by the source localization

results showing an increase of amplitude in left frontal area for the second presentation ([Figure S3](#)). However, the phase precession observed here is unlikely to be accounted by mechanisms solely related to memory, which is often shown within a higher frequency band (i.e., the theta band, 4–8 Hz) [[24](#)–[27](#)]. The fact that the phase precession only happened in the left hemisphere and at the frequency corresponding to the line rate echoes the finding in [Figure 2E](#) and demonstrates that the phase precession is primarily involved in the process of segmenting line structures of *Jueju*.

In summary, these results further demonstrate that speech segmentation is driven by listeners' prior knowledge of not only formality of speech materials but also speech contents [[28](#)]. Speech segmentation involves predictive processes as a result of listeners' prior knowledge of speech materials.

### Word Co-occurrence Negatively Correlates with Speech Segmentation

It has been shown that statistical regularity of linguistic elements aids in speech segmentation [[5](#), [9](#)]. Here, we quantified statistical co-occurrence of monosyllabic words in the poem materials using an index of language perplexity, and examined whether the segmentation of poetic speech stream correlated with the language perplexity. The language perplexity of a poem is the inverse probability of the test lines, normalized by the number of words [[29](#)] ([STAR Methods](#)) and represents how likely words of a poem tend to appear together in the language corpus used



**Figure 4. Correlation between Language Perplexity and Speech Segmentation**

(A) Distribution of the language perplexity. The language perplexity indicates uncertainty of words in the probability distribution of poetic contents derived from the training corpus.

(B) Language perplexity and neural correlations. The line color codes hemisphere and presentation. The transparent thin lines (same color code) represent the upper and lower thresholds derived from a permutation test ( $p < 0.01$ ). The significant correlations were observed for the first presentation in the left hemisphere from 0.66 to 0.8 Hz. The shaded areas represent one standard error of the mean.

(C) Topographies of language perplexity and neural correlations from 0.66 to 0.8 Hz. See also [Figure S4](#).

for training the neural network. The language perplexities of all the poems are shown in [Figure 4A](#). We correlated the language perplexity with the neural responses to the poem stimuli from the ten channels selected at  $\sim 0.24$  and  $\sim 0.64$  Hz, respectively ([Figures 4B](#) and [S4](#)). We conducted a permutation test by shuffling the poem tokens and derived thresholds of the two-sided alpha level of 0.01 ([STAR Methods](#)). We chose a threshold of 0.01 instead of 0.05 because four conditions (two hemispheres and two presentations) were tested. We observed a significant positive correlation between neural responses from 0.66 to 0.8 Hz in the left hemisphere and the language perplexity during the first presentation ([Figure 4B](#)), which echoes the finding on the line parsing ([Figure 2E](#)). We further calculated the correlations over all the channels and show the topography in [Figure 4C](#). The positive correlation was only significant for the first presentation, probably because the representation of the line structure was established during the first presentation and the brain relied less on the external features of the poems to process the poem stimuli during the second presentation. It is worth noting that the lower the language perplexity of a poem is, the higher the statistical co-occurrence is among words. Therefore, this finding shows that the word co-occurrence, quantified by the language perplexity, negatively correlated with speech segmentation.

One explanation for the negative correlation between the statistical co-occurrence of monosyllabic words and the line segmentation is that, when the language perplexity of a poem is low, words and phrases within the poem are seemingly related to each other within and across lines. This smeared boundaries between lines and conflicted with listeners' knowledge of the poetic structure; worse speech segmentation around the line rate was observed. In contrast, the poems of high language perplexity have words and phrases that do not usually appear together in the representative language corpus used for training the neural network, which made each word or phrase stand out, so that listeners can better group those independent words into lines based on the poetic structure. Therefore, better line segmentation was observed. Complementary to the previous finding of the statistical regularity aiding in

speech segmentation [[5](#), [9](#)], our finding suggests that the segmentation cues in a speech stream are selectively deployed by the brain according to cognitive tasks and contexts. In the current context of grouping poetic languages that deviate from ordinary language usage, listeners segmented the poetic speech stream according to their prior knowledge of poetic structure while assigning less weight to the statistical cues.

In summary, we investigated how the poetic structures aid speech segmentation and how listeners' prior knowledge of the poetic structures contribute to speech segmentation. We employed artificial intelligence and neurophysiology collaboratively and gained insights from the neural correlates of speech segmentation—we observed a multi-timescale process of active speech segmentation driven by listeners' prior knowledge of the formal and conceptual structure of poems ([Figure 2](#)). Unprecedentedly, we found the phenomenon of phase precession in speech perception due to repetition of poems, reflecting a predictive process of speech segmentation ([Figure 3](#)). The statistical co-occurrence of monosyllabic words in the poems did not explain this structure-based speech segmentation ([Figure 4](#)). Our findings demonstrate that, in addition to grammatical structures, semantic cohorts, and statistical cues, listeners' prior knowledge of the structural regularity of language materials can be exploited to actively segment speech streams and group linguistic contents.

## STAR★METHODS

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

- [KEY RESOURCES TABLE](#)
- [LEAD CONTACT AND MATERIALS AVAILABILITY](#)
- [EXPERIMENTAL MODEL AND SUBJECT DETAILS](#)
- [METHOD DETAILS](#)
  - Artificial poem generation
  - MEG measurement
  - Experimental procedure
  - MEG recording and preprocessing

## ● QUANTIFICATION AND STATISTICAL ANALYSIS

- Modulation spectra of the poem stimuli
- MEG Spectral analysis
- Acoustic-neuro coherence
- MEG Temporal analyses
- MEG source reconstruction
- ROI analysis of neural dynamics in the source space
- Correlations between neural signals and language perplexity
- Simulation on examining the origin of the effect at ~0.24 Hz

## ● DATA AND CODE AVAILABILITY

## ● ADDITIONAL RESOURCES

### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cub.2020.01.059>.

A video abstract is available at <https://doi.org/10.1016/j.cub.2020.01.059#mmc3>.

### ACKNOWLEDGMENTS

We thank Jeff Walker and Jess Rowland for their technical support, and David Poeppel and Winfried Menninghaus for their comments on previous versions of the manuscript. This study was supported by the National Natural Science Foundation of China 31871131 to X. Tian, and 31970987 and 31771210 to Q.C.; Major Program of Science and Technology Commission of Shanghai Municipality (STCSM) 17JC1404104; the Program of Introducing Talents of Discipline to Universities, Base B16018 to X. Tian; the JRI Seed Grants for Research Collaboration from NYU-ECNU Institute of Brain and Cognitive Science at NYU Shanghai to X. Tian and Q.C.; and NIH 5R01DC005660 to David Poeppel and the Max Planck Society.

### AUTHOR CONTRIBUTIONS

X.Teng and X.Tian designed the study. X. Teng performed the experiment and analyzed all the data; M.M. generated poem materials; J.Y. and Q.C. conducted online evaluations; X. Teng, S.B., and X. Tian interpreted data; X. Teng and X. Tian wrote the first draft of the paper; and X. Teng, M.M., J.Y., S.B., Q.C., and X. Tian wrote the paper. X. Tian supervised the project.

### DECLARATION OF INTERESTS

The authors declare no competing financial interests.

Received: July 10, 2019

Revised: November 8, 2019

Accepted: January 17, 2020

Published: March 5, 2020

### REFERENCES

1. Egan, C.H. (1993). A Critical Study of the Origins of 'Chüeh-chü' Poetry. *Asia Major* 6, 83–125.
2. (2002). 王, 力. 诗词格律概要 (北京出版社).
3. Cai, Z.-Q. (2008). How to read Chinese poetry: A guided anthology (Columbia University Press).
4. Steinhauer, K., and Friederici, A.D. (2001). Prosodic boundaries, comma rules, and brain responses: the closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *J. Psycholinguist. Res.* 30, 267–295.
5. Sanders, L.D., Newport, E.L., and Neville, H.J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nat. Neurosci.* 5, 700–703.
6. Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., and Pallier, C. (2015). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron* 88, 2–19.
7. Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158–164.
8. Goucha, T., and Friederici, A.D. (2015). The language skeleton after dissecting meaning: A functional segregation within Broca's Area. *Neuroimage* 114, 294–302.
9. Saffran, J.R., Aslin, R.N., and Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.
10. Fletcher, W.J.B. (1966). *Gems of Chinese Verse, and More Gems of Chinese Poetry* (Paragon Book Reprint Corp.).
11. Dehaene-Lambertz, G., Dehaene, S., Anton, J.L., Campagne, A., Ciuciu, P., Dehaene, G.P., Degenhien, I., Jobert, A., Lebihan, D., Sigman, M., et al. (2006). Functional segregation of cortical language areas by sentence repetition. *Hum. Brain Mapp.* 27, 360–371.
12. Summerfield, C., Wyart, V., Johnen, V.M., and de Gardelle, V. (2011). Human Scalp Electroencephalography Reveals that Repetition Suppression Varies with Expectation. *Front. Hum. Neurosci.* 5, 67.
13. Friederici, A.D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84.
14. Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
15. Pallier, C., Devauchelle, A.-D., and Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci. USA* 108, 2522–2527.
16. Scott, S.K., and Johnsrude, I.S. (2003). The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* 26, 100–107.
17. Zatorre, R.J., Belin, P., and Penhune, V.B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46.
18. Zhang, X., and Lapata, M. (2014). Chinese poetry generation with recurrent neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 670–680.
19. Steinhauer, K., Alter, K., and Friederici, A.D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat. Neurosci.* 2, 191–196.
20. Chu, H.-C.J., Swaffar, J., and Charney, D.H. (2002). Cultural representations of rhetorical conventions: The effects on reading recall. *TESOL Quarterly* 36, 511–542.
21. Buzsáki, G. (2005). Theta rhythm of navigation: link between path integration and landmark navigation, episodic and semantic memory. *Hippocampus* 15, 827–840.
22. Lisman, J. (2005). The theta/gamma discrete phase code occurring during the hippocampal phase precession may be a more general brain coding scheme. *Hippocampus* 15, 913–922.
23. Jensen, O., and Lisman, J.E. (1996). Hippocampal CA3 region predicts memory sequences: accounting for the phase precession of place cells. *Learn. Mem.* 3, 279–287.
24. Heusser, A.C., Poeppel, D., Ezzyat, Y., and Davachi, L. (2016). Episodic sequence memory is supported by a theta-gamma phase code. *Nat. Neurosci.* 19, 1374–1380.
25. Siegel, M., Warden, M.R., and Miller, E.K. (2009). Phase-dependent neuronal coding of objects in short-term memory. *Proc. Natl. Acad. Sci. USA* 106, 21341–21346.
26. Griffiths, B.J., Parish, G., Roux, F., Michelmann, S., van der Plas, M., Kolibius, L.D., Chelvarajah, R., Rollings, D.T., Sawlani, V., Hamer, H., et al. (2019). Directional coupling of slow and fast hippocampal gamma with neocortical alpha/beta oscillations in human episodic memory. *Proc. Natl. Acad. Sci. USA* 116, 21834–21842.
27. Hanslmayr, S., Axmacher, N., and Inman, C.S. (2019). Modulating Human Memory via Entrainment of Brain Oscillations. *Trends Neurosci.* 42, 485–499.

28. Cervantes Constantino, F., and Simon, J.Z. (2018). Restoration and Efficiency of the Neural Processing of Continuous Speech Are Promoted by Prior Knowledge. *Front. Syst. Neurosci.* *12*, <https://doi.org/10.3389/fnsys.2018.00056>.
29. Jurafsky, D., and Martin, J.H. (2014). *Speech and language processing Volume 3* (Pearson London).
30. He, J., Zhou, M., and Jiang, L. (2012). Generating chinese classical poems with statistical machine translation models. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.
31. Yan, R., Jiang, H., Lapata, M., and Lin, S. (2013). I, poet: automatic chinese poetry composition through a generative summarization framework under constrained optimization. In *Twenty-Third International Joint Conference on Artificial Intelligence*.
32. Roberts, T.P., Ferrari, P., Stufflebeam, S.M., and Poeppel, D. (2000). Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception. *J. Clin. Neurophysiol.* *17*, 114–129.
33. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* *2011*, 156869.
34. de Cheveigné, A., and Simon, J.Z. (2007). Denoising based on time-shift PCA. *J. Neurosci. Methods* *165*, 297–305.
35. de Cheveigné, A., and Simon, J.Z. (2008). Sensor noise suppression. *J. Neurosci. Methods* *168*, 195–202.
36. Ellis, D.P.W. (2009). Gammatone-like spectrograms. <http://www.ee.columbia.edu/~dpwe/resources/matlab/gammatonegram/>.
37. Patterson, R.D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (1987). An efficient auditory filterbank based on the gammatone function. In a meeting of the IOC Speech Group on Auditory Modelling at RSRE 2 (Institute of Acoustics on Auditory Modelling).
38. Glasberg, B.R., and Moore, B.C.J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* *47*, 103–138.
39. Søndergaard, P.L., and Majdak, P. (2013). In *The Technology of Binaural Listening*, J. Blauert, ed. (Springer Berlin Heidelberg), pp. 33–56.
40. Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* *164*, 177–190.
41. Peelle, J.E., Gross, J., and Davis, M.H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* *23*, 1378–1387.
42. Doelling, K.B., Arnal, L.H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage* *85*, 761–768.
43. Kay, S.M. (1988). *Modern spectral estimation* (Pearson Education India).
44. Berens, P. (2009). CircStat: a MATLAB toolbox for circular statistics. *J. Stat. Softw.* *31*, <https://doi.org/10.18637/jss.v031.i10>.
45. Gramfort, A., et al. (2014). MNE software for processing MEG and EEG data. *NeuroImage* *86*, 446–460.
46. Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature* *536*, 171–178.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE   | SOURCE   | IDENTIFIER   |
|---|--|--|
| Deposited Data  |  |  |
| Texts and sound files of poems                              | Github   | <a href="https://ray306.github.io/neural_poem/poems.html">https://ray306.github.io/neural_poem/poems.html</a>  |
| MEG data  | OSF data repository  | <a href="https://osf.io/jg3xh/">https://osf.io/jg3xh/</a>  |
| Online evaluation data                                      | Github   | <a href="https://ray306.github.io/neural_poem/beh_rating.html">https://ray306.github.io/neural_poem/beh_rating.html</a>  |
| Software and Algorithms                                     |  |  |
| Presentation Software                                       | Psychtoolbox-3   | <a href="http://psychtoolbox.org/">http://psychtoolbox.org/</a> ; RRID: SCR_002881   |
| MATLAB  | The Mathworks  | RRID: SCR_001622   |
| Python 3.7.1  | Anaconda   | <a href="https://www.anaconda.com/">https://www.anaconda.com/</a> ; RRID: SCR_008394   |
| MNE-Python  | MNE Developers ( <a href="https://mne.tools/">https://mne.tools/</a> ) | RRID: SCR_005972   |
| Fieldtrip 20181014  | Donders Institute for Brain, Cognition and Behavior                    | <a href="http://www.fieldtriptoolbox.org/">http://www.fieldtriptoolbox.org/</a> ; RRID: SCR_004849   |
| Neospeech synthesizer                                       | ReadSpeaker  | <a href="http://www.neospeech.com/">http://www.neospeech.com/</a>  |
| Circular Statistics Toolbox                                 | Philipp Berens   | <a href="https://de.mathworks.com/matlabcentral/fileexchange/10676-circular-statistics-toolbox-directional-statistics">https://de.mathworks.com/matlabcentral/fileexchange/10676-circular-statistics-toolbox-directional-statistics</a> ; RRID: SCR_016651 |
| GammatoneFast   | D. P. W. Ellis   | <a href="http://www.ee.columbia.edu/~dpwe/resources/matlab/gammatonegram/">http://www.ee.columbia.edu/~dpwe/resources/matlab/gammatonegram/</a>  |
| sqdDenoise  | Jonathan Z. Simon, Alain de Cheveigné                                  | <a href="http://cansl.isr.umd.edu/simonlab/Resources.html">http://cansl.isr.umd.edu/simonlab/Resources.html</a> ; <a href="http://audition.ens.fr/adc/NoiseTools/">http://audition.ens.fr/adc/NoiseTools/</a>  |
| Chinese Poetry Generation with Recurrent Neural Networks    | Github   | <a href="https://github.com/XingxingZhang/rnnpg">https://github.com/XingxingZhang/rnnpg</a>  |
| Other   |  |  |
| 157 axial gradiometer whole head MEG at New York University | KIT, Kanazawa, Japan   | N/A  |

### LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Xing Tian ([xing.tian@nyu.edu](mailto:xing.tian@nyu.edu)). This study did not generate new unique reagents.

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

Thirteen Chinese native speakers studying at New York University participated in the MEG experiment (6 females; age ranging from 23 to 30; 12 right-handed and 1 left-handed). None of the participants majored in Chinese literature; all of the participants were fully familiar with the structure of Chinese ancient poetry and acquired their knowledge of *Jueju* from education acquired before college. None of the participants had any prior experience on any of the artificial poems before the magnetoencephalography experiment. All the participants had normal hearing and reported no neurological deficits according to their self-report. Written consent forms were obtained before the experiment. The study was approved by the New York University Institutional Review Board (IRB# 10–7277) and conducted in conformity with the 45 Code of Federal Regulations (CFR) part 46 and the principles of the Belmont Report.

### METHOD DETAILS

#### Artificial poem generation

We replicated the poem generation procedures in Zhang and Lapata (2014) ([github.com/XingxingZhang/rnnpg](https://github.com/XingxingZhang/rnnpg)) and constructed a poem generation model using recurrent neural networks. We brief the procedures here. The model primarily selected poetic contents (words and lines) constrained by the learned surface realization (e.g., phonological rules). Trained on a corpus of classical Chinese poems that contained 78,859 quatrains from Tang Poems, Song Poems, Song Ci, Ming Poems, Qing Poems, and Tai Poems, the model learned representations of individual characters and their combinations into one or more lines, as well as how these lines

mutually reinforce and constrain each other. The loss function was determined by the cross-entropy between the predicted character distribution and the actual character distribution in our corpus. An L2 regularization term was employed. During poem generation, some keywords (e.g., spring, lute, drunk) were first supplied, which defined the main topic of a poem, and then the generation model expanded the keywords into a set of phrases which were restricted to the *ShiXueHanYing* poetic phrase taxonomy [30, 31]. From there, the model produced the first line of the poem based on these keywords. Subsequent lines were generated based on all previous lines, which were subject to phonological (e.g., phonological rules) and structural constraints (the length of each line). An illustration of the poem generation is shown in Figure 1B, which was adapted from Zhang and Lapata, 2014.

We first used the poem model to generate 4000 artificial poems. To select artificial poems within a representative range of qualities, one of the authors (M.M), who had expertise on Chinese ancient poetry, rated the generated poems and manually grouped them into three categories – ‘good’, ‘medium’, and ‘bad’. From these three categories, we randomly selected 32 ‘good’ poems, 86 ‘medium’ poems, and 32 ‘bad’ poems to constitute 150 artificial poems used in the MEG experiment. We further acquired an objective index of complexity of these 150 artificial poems - *language perplexity*. The language perplexity gave a convenient intuition of the degree of uncertainty in the probability distribution of poetic contents derived from the training corpus - the lower the language perplexity is, the more likely the corresponding poem is generated by our language model trained on the poetry corpus. The words in the poem of low language perplexity tend to co-occur in the poetry corpus.

In computational linguistics, a general way to evaluate the performance of a trained language model is to calculate perplexity on held-out lines. The language perplexity is the inverse probability assigned to a held-out line, normalized by the number of words it contains [29]. It is equivalent to 2 to the power of cross-entropy between the true probabilistic distribution of lines in the training corpus and the actual probabilistic distribution of the held-out lines. If our trained neural network language model had learned the probabilistic distribution of poetic contents in the corpus and the held-out lines were drawn from an unknown probabilistic distribution, the cross-entropy quantifies the realized average encoding size (a measurement of uncertainty) using the distribution of poetic contents of the training corpus and the actual encoding size based on the probabilistic distribution from which the held-out lines were drawn. Here, we calculated the language perplexity as below: the language perplexity of a poem is the inverse probability of the test lines ( $w$ ), normalized by the number of words ( $N$ ):

$$P(W) = P(w_1, w_2, \dots, w_N)^{-1/N}$$

According to the chain rule of probability, we derived:

$$P(W) = \left[ \prod_{i=1}^N \frac{1}{P(w_i | w_1, w_2, \dots, w_{i-1})} \right]^{-1/N}$$

A recurrent neural network model was used to compute the conditional probability

$$P(w_i | w_1, w_2, \dots, w_{i-1})$$

We further selected 30 poems written by humans and evaluated the language perplexity to make a comparison between artificial poems and real poems. The 30 real poems were further used in the MEG experiment and were less known to Chinese listeners, so that they matched the familiarity of the artificial poems. The histogram of the language perplexity of all the selected poems is shown in Figure 4A.

### MEG measurement

**Poem stimuli.** The 150 artificial poems and the 30 human written poems were included in the MEG experiment. We used the same procedure as in the study of Ding et al. [7] to synthesize poem stimuli. All monosyllabic characters in the poems were synthesized individually using the Neospeech synthesizer (<http://www.neospeech.com/>, the male voice, Liang). The synthesized syllables were adjusted to 300 ms by truncation or padding silence at the end. If a syllable was truncated, the last 25 ms of this syllable was smoothed by a cosine window. Each poem was constructed by concatenating the waveforms of characters, based on the poem structure, into an isochronous sequence of syllables without inserting any acoustic gaps. By following this procedure, we eliminated any acoustic cues, such as co-articulation and intonation in poems recited by humans, so that we can investigate how poems are parsed by listeners according to their forms and structures. Each poem contained 20 syllables and four lines. Therefore, each poem stimulus was 6 s long with the syllable rate around 3.33 Hz, the line rate was around 0.67 Hz, and the rhyming rate was around 0.33 Hz (Figure 1C).

All the poem stimuli were normalized to ~65 dB SPL. The stimuli were delivered through plastic air tubes connected to foam earpieces (E-A-R Tone Gold 3A Insert earphones, Aearo Technologies Auditory Systems) in the experiment.

### Experimental procedure

Before the MEG recording, the participants were required to read all the selected 180 poems within 5 - 10 min. They were asked to answer the question ‘how good is the poem’ instructed in Chinese and rated the poems from 1 (bad) to 3 (good). This procedure aimed to familiarize the participants with the poem materials but prevented them from memorizing the poems as the time period was too short for the participants to memorize all the 180 poems. The 180 poems were displayed in four Excel sheets and the order of sheets for rating was counter-balanced among the participants.

During the MEG recording, participants were presented each poem stimulus twice and rated the poems from 1 (bad) to 3 (good) after the second presentation by pressing one of three buttons using their right hand. On each trial, participants were first required to focus on a white fixation in the center of a black screen. After 2.5 ~3.5 s, the fixation disappeared and one poem stimulus was presented. After the poem stimulus was over, the white fixation reappeared. The same poem was presented again 1 s later. After the second presentation, a question in Chinese, ‘how good is the poem?’, was displayed on the screen and the participants were required to rate the poem. The next trial started in 2.5 ~3.5 s after the button press. The procedure was illustrated in [Figure 1C](#). The poem stimuli were presented in three blocks and the order of the poems was randomized for each participant.

### MEG recording and preprocessing

MEG signals were recorded with participants in a supine position, in a magnetically shielded room using a 157-channel whole-head axial gradiometer system (KIT, Kanazawa Institute of Technology, Japan). The MEG data were acquired with a sampling rate of 1,000 Hz, filtered online with a low-pass filter of 200 Hz, with a notch filter at 60 Hz. After the main experiment, participants were presented with 1 kHz tone beeps of 50 ms duration as a localizer to determine their M100 evoked responses, which is a canonical auditory response [32]. Twenty channels with the largest M100 response in both hemispheres (10 channels in each hemisphere) were selected as auditory channels for each participant individually.

MEG data analyses in sensor space were conducted in MATLAB using the Fieldtrip toolbox 20181014 [33]. Raw MEG data were noise-reduced offline using the time-shifted PCA [34] and sensor noise suppression [35] methods. The data of each block were high-pass filtered post-acquisition at 0.1 Hz using a two-pass third-order Butterworth filter. Each presentation of the poem stimuli in a trial was extracted into a 11 s epoch, with 3 s pre-stimulus period and 2 s post-stimulus period. Each epoch went through a low-pass two-pass fourth-order Butterworth filter at 60 Hz (the default setting of Fieldtrip toolbox).

All trials were visually inspected, and those with artifacts such as channel jumps and large fluctuations were discarded. The baseline was corrected for each trial by subtracting out the mean of the whole trial before conducting further analyses. An independent component analysis was used to correct for eye blink-, eye movement-, heartbeat-related and system-related artifacts. As recordings from some participants were very noisy, up to 30 trials were excluded for a few participants. To keep the number of trials consistent across the participants and between the presentations, we kept 150 trials from the first presentation and from the second presentation, respectively, for each participant. The removed trials were randomly selected and we did not take into consideration whether a removed trial was from an artificial poem or a real one. As both the artificial poems and the real poems were novel to the participants, we grouped all the poems to increase statistical power. The 150 trials of each presentation included both the artificial poems and the real poems.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Modulation spectra of the poem stimuli

To characterize acoustic properties of the synthesized poem stimuli, we derived amplitude envelopes of the poem stimuli by simulating outputs of cochlear filters using a gammatone filterbank. We filtered the audio files through a gammatone filterbank of 64 bands logarithmically spanning from 80 Hz to 6400 Hz [36, 37]. The amplitude envelope of each cochlear band was extracted by applying the Hilbert transformation on each band and taking the absolute values [38, 39]. The amplitude envelopes of 64 bands were then averaged and down-sampled from the sampling rate of 16000 Hz to 1000 Hz to match the sampling rate of the processed MEG signals for further analyses (i.e., acoustic-neuro coherence).

We transformed the amplitude envelopes to modulation spectra so that the canonical temporal dynamics of the poem stimuli can be concisely shown in the spectral domain. We first calculated the modulation spectrum of the amplitude envelope of each poem stimuli using the FFT and took the absolute value of each frequency point. We then averaged the modulation spectra across the 180 poem stimuli and showed the averaged modulation spectrum in [Figure 1D](#).

### MEG Spectral analysis

We first conducted spectral analyses using the auditory channels independently selected for each participant using the auditory localizer, as we assumed for this first step of analysis that the auditory stimuli would evoke strong auditory-related responses, which should be mostly reflected in the auditory channels. We first averaged in the temporal domain over all the trials selected after preprocessing from both the first and the second presentations, so that we had enough analytic power to extract neural signatures involved with parsing the poem stimuli. On each channel, the averaged data from 0.5 s pre-stimulus period and 6.5 s post-stimulus period were selected and a Hanning window of 7 s was applied on the selected data. We converted the data from the temporal domain to the spectral domain using the fast Fourier transform (FFT) and derived amplitude spectra by taking the absolute value of each complex number. The length of data points was 7000 and we padded 43000 data points so that the spectral resolution of the amplitude spectra was 0.02 Hz, which guaranteed that we had sufficient spectral resolution to observe neural signatures around the rhyming rate (~0.33 Hz) or the line rate (~0.67 Hz). We averaged the amplitude spectra over all the auditory channels and derived an averaged amplitude spectrum for each participant.

To remove the influence of the 1/f shape of the spectrum and to reveal neural signatures in the frequency range below 1 Hz, we employed a surrogate procedure. For each presentation in a trial, we randomly selected a new onset point between 2.5 pre-stimulus period and 1 s post-stimulus period and defined a new data series of 6.5 s. We conducted the spectral analysis to derive an amplitude

spectrum for each participant and calculated a group-averaged amplitude spectrum from this new dataset. This procedure was repeated 1000 times and we sorted the results at each frequency point to derive a threshold of a one-sided alpha level of 0.01. By subtracting the mean of the amplitude distribution of the surrogate procedure from the empirical amplitude spectrum, the neural components locked to the poetic structure were revealed (Figure 2A, lower panel), as the surrogate procedure disrupted the temporal correspondence between the poem stimuli and the neural signals.

The results above revealed two spectral components ( $\sim 0.24$  Hz and  $\sim 0.64$  Hz):  $\sim 0.64$  Hz was potentially relevant to parsing line structures of the poem stimuli while  $\sim 0.24$  Hz indicated existence of a neural component at a timescale above the line structure. We next selected MEG channels that showed the highest amplitude of the two components based on the group-averaged results (Figures 2A and 2B) for further investigation on lateralization and the effects of repetition of the poem stimuli. We selected ten channels in each hemisphere and averaged the power spectra over the channels in each hemisphere for each participant. The number of the channels selected was arbitrarily determined, so that the decision could affect our results. We varied the channel selection and show that the channel selection did not affect the main findings qualitatively nor the conclusion of the study (Figure S2).

To determine the significant frequency ranges showing differences between hemispheres and presentations, we first conducted a cluster-based permutation test from 0.1 to 2 Hz on the four combinations of two factors - hemisphere (2 levels) and presentation order (2 levels) [40]. We shuffled the labels of the four combinations in each participant's data and created a new group dataset. Then, we conducted a one-way rmANOVA at each frequency point. We set the alpha level of the threshold as 0.05 following the previous conventions and grouped frequency points above this threshold into clusters. We calculated the cluster statistics (F value) and kept the cluster of the highest F value. This procedure was repeated 1000 times and we sorted the F values over 1000 clusters. The 99 percentile was chosen as the threshold.

### Acoustic-neuro coherence

We calculated the coherence between the neural signals and the amplitude envelopes of the poem stimuli in the spectral domain. The rationale here is that, if certain acoustic properties in the poem stimuli indicate boundaries of lines or other poetic structures, high coherence between the acoustics and the neural signals should be observed at the corresponding timescales of the poetic structures. If such coherence is not observed, it can be concluded that listeners do not rely on the acoustic cues to segment poetic streams. This measurement has been used in neurophysiological studies on speech perception [41, 42], which quantified how neural responses evoked by auditory stimuli correlate with acoustic dynamics.

We here calculated magnitude-squared coherence [43] using the function 'mscohere' in MATLAB 2016b (<https://de.mathworks.com/help/signal/ref/mscohere.html>). The calculation included the neural signal of each trial from 0 s to 6 s, which matched the length of the poem stimuli. We zero-padded 50000 points to ensure enough spectral resolution. No windowing function was used so that the coherence was calculated on the entire poem stimuli. To determine the significant frequency ranges showing differences between hemispheres and presentations, we first conducted a cluster-based permutation test from 0.1 to 7 Hz on the four combinations of two factors - hemisphere (2 levels) and presentation order (2 levels). Here, we selected an upper bound of 7 Hz because we aimed to include the first harmonic component of the syllable rate (6.67 Hz). The procedure here was the same as in *MEG Spectral analysis* above.

### MEG Temporal analyses

To show temporal dynamics of the neural responses reflecting parsing poetic structures of the poem stimuli (Figure 3A), we filtered each trial using a low-pass two-pass third-order Butterworth filter at 2 Hz to remove neural responses to acoustic properties of the poem stimuli (Figure 1C).

To investigate the temporal dynamics of the two frequency components ( $\sim 0.24$  Hz and  $\sim 0.64$  Hz), we selected two frequency ranges: 0.1 – 0.4 Hz and 0.4 – 0.96 Hz. The rationale was that a frequency component in neural signals is often not a strictly sinusoidal component but a narrowband component with a center frequency. The spectral ranges of an amplitude peak in the spectrum reflect the bandwidth of the narrowband frequency component. If we only selected a part of this peak range, this would distort the temporal characteristics of this frequency component. The rationale for selecting the 0.1 – 0.4 Hz range was that we first constrained the low spectral bound to be 0.1 Hz, as the cutoff frequency of the high pass filter applied in the preprocessing step was 0.1 Hz. Then, we averaged the amplitude spectra over two hemispheres and two presentations, so that a group-averaged amplitude spectrum was derived, which was unbiased to hemispheres and presentations. Based on this amplitude spectrum, we selected 0.4 Hz, which was the trough between the spectral peaks of the two frequency components, detected by calculating the second derivative of the spectrum. For the 0.4 - 0.96 Hz range, we selected the frequency point of 0.96 Hz, which was the upper bound of the spectral peak of the component of  $\sim 0.64$  Hz. We did not select the frequency ranges where significant effects of hemisphere and presentation order were found, because the significant ranges did not cover the whole spectral peaks of the two frequency components.

We applied a bandpass two-pass third-order Butterworth filter on each trial of the preprocessed data to extract temporal signals in both the frequency ranges. We down-sampled the data from 1000 Hz to 20 Hz, as the sampling rate of 20 Hz was enough to carry spectral components in the filtered data and fast calculations could be conducted. For each participant, we averaged over 150 trials of each presentation and then applied the Hilbert transform on the averaged data. The phase series was extracted from the complex number of each time point. We calculated the differences of phase between the two presentations and then derived circular means and circular standard deviations over the participants (Figure 3C). We conducted a one-sample circular t test for the mean phase at each time point from –1 s to 7 s and set the alpha level to be 0.01 by assuming the mean direction to be zero without binning the phase

data [44]. This circular test calculated confidence intervals based on the chosen alpha level and determined whether the phase difference between the second and the first presentations was larger than the confidence interval – at which time point the phase difference significantly differed from zero. We chose a stringent alpha level of 0.01 in order to decrease the type I error, as four conditions were tested here (two hemispheres and two presentations).

### MEG source reconstruction

Each participant's head shape was digitized (Polhemus 3SPACE FASTRAK) before the MEG recording session. Five marker coils were attached to their head and were localized with respect to the MEG sensors at the beginning and at the end of the MEG recording. We generated individual head models by co-registering the FreeSurfer average brain (CorTechs Labs Inc., Lajolla, CA) based on individual head shapes and position, and by using uniform scaling, translation, and rotation in MNE-Python [45]. We did not use participants' MRI anatomy of the brain for generating head models. The source reconstructions were done by estimating the cortically constrained dynamic statistical parametric mapping (dSPM) of the MEG data. The forward solution (magnetic field estimates at each MEG sensor) was estimated from a source space of 5121 activity points with a boundary-element model (BEM) method. The inverse solution was calculated from the forward solution. Subsequently, we morphed each individual brain back to the FreeSurfer average brain to obtain the group-averaged results.

We matched the parameters used in Fieldtrip during preprocessing and conducted preprocessing in MNE-python (lowpass filtering at 60 Hz and independent component analysis). After preprocessing, we derived single-trial data in source space, which were then exported to MATLAB. We conducted the same spectral analysis in the source space as that in the sensor space in MATLAB and converted the data back to Python format to create the source plots of each frequency components (Figure 2). Similarly, we repeated the temporal analysis in the source space and generated the source plots (Figure S3A).

### ROI analysis of neural dynamics in the source space

Regions of interests (ROI) in the frontal and temporal lobes were defined based on anatomical labels in the HCP-MMP1.0 parcellation [46]. For the frontal ROI, four areas ('FOP5', '45', 'p47r', and 'IFSa' as labeled in the parcellation) were included. For the temporal ROI, three areas ('STGa', 'TA2', and 'PI' as labeled in the parcellation) were included. The ROI selections aimed to match the areas showing increased neural activity in the second presentation (Figure S3A). This selection resulted in four ROIs, one frontal ROI and one temporal ROI in each hemisphere (Figure S3B). We took absolute values of each source dipole and derived a mean over all the source dipoles in each ROI.

### Correlations between neural signals and language perplexity

We correlated the amplitude spectra from 0.05 Hz to 5 Hz of the neural signals with the language perplexity to investigate whether the language perplexity modulated the neural signatures that reflected poem parsing processes. The correlation was conducted using the channels selected based on the two frequency components, respectively (Figure 2D, inserts).

For the correlation between the language perplexity and the neural signals, we included both the artificial poems and the real poems. At each frequency point, we correlated the amplitude values of the trials from the artificial poems with their corresponding language perplexity using Pearson correlation. We applied Fisher-z transformation on the correlation coefficients of each participant and averaged the transferred coefficients over all the participants. To determine the significance of correlations, we conducted a permutation test. We randomly shuffled the correspondence between the trials and the poem stimuli and derived a correlation coefficient from the shuffled dataset for each participant. We repeated the group-averaging procedure and derived a new group-averaged coefficient. This procedure was repeated 1000 times and we sorted the results at each frequency point to derive a threshold of a two-sided alpha level of 0.01 (the thin lines in Figures 4B and S4A), as four conditions were tested.

### Simulation on examining the origin of the effect at $\sim 0.24$ Hz

Our focus on the neural signals within the frequency range around 0.24 Hz was motivated from two observations: 1) after subtracting out the baseline derived from the surrogate test (*MEG Spectral analysis*), we observed a peak at  $\sim 0.24$  Hz in the amplitude spectrum (Figure 2A, lower panel); 2) correspondingly, a repetition effect was found between 0.12 Hz and 0.28 Hz (Figure 2D, upper panel). The spectral component of  $\sim 0.24$  Hz did not sufficiently suggest existence of rhythms at  $\sim 0.24$  Hz, as the signal length (7 s) used to calculate the amplitude spectra was less than two cycles of  $\sim 0.24$  Hz. It is helpful to further examine the origin of the effect at  $\sim 0.24$  Hz, so that a proper explanation can be provided. Therefore, we conducted a series of simulations and related the simulations to the data to further reveal the origin of the effect at  $\sim 0.24$  Hz. In short, we showed that the spectral component of  $\sim 0.24$  Hz represented a combination of the onset responses to the poem stimuli and a downward amplitude modulation from the first line to the third line in *Jueju*, and that the repetition effect around 0.24 Hz (Figure 2D, upper panel) originated from the decreased amplitude modulation on the poetic line segmentation from the first presentation to the second presentations.

The temporal range ( $-0.5$  s to  $6.5$  s) used to calculate the amplitude spectra for the results in Figure 2D included four major components – the onset response, the offset response, the neural signals corresponding to the poetic line segmentation, and a downward amplitude modulation on the line segmentation (Figure S2C). One or a few of these four major components probably contributed to the effect at  $\sim 0.24$  Hz, so we constructed models to examine the effect of each of the four components on the spectral component of  $\sim 0.24$  Hz. Admittedly, the neural dynamics in the data is more complex and our abstraction may not fully grasp the details of the neural signals. For example, we did not include in the models the phase precession (Figure 3C). However, the following simulations

showed that these four major components were sufficient to explain the origin of the effect at  $\sim 0.24$  Hz. The MATLAB codes used to conduct the simulation analyses and to plot the corresponding figures for illustration were deposited in the OSF project folder, and we also deposited the figures generated from the simulation analyses.

We constructed four models based on our above abstraction to capture the neural dynamics in Figure 3A (MATLAB script: *Simulation\_1.m*). The model outputs in the temporal and spectral domains are shown in Figure S2D. Model 1 included a component of the line segmentation. The neural signals corresponding to the line segmentation were simulated by a sinusoid wave at  $\sim 0.67$  Hz (the line rate). To simplify the model, we assumed that the peaks of the sinusoid wave aligned perfectly with the line boundaries. The onset of the simulated signal for the line segmentation started from 750 ms, the middle point of the first line. Therefore, the sinusoid wave included three cycles, each of which corresponded to one line. Following the procedures used to calculate the amplitude spectra in Figure 2D, we calculated the amplitude spectrum of the simulated line segmentation. The free parameter of Model 1 is the amplitude of the line segmentation and was set to 1, initially.

Model 2 was constructed based on Model 1 and the amplitude modulation from the first line to the third line was added. The amplitude modulation was hypothesized or abstracted to be a straight line with its slope as a free parameter. The modulation effect was imposed on the line segmentation from 750 ms to 5250 ms. The intersection of the modulation line was not fixed. Alternatively, we imposed a constraint driven by the results at  $\sim 0.64$  Hz (Figures 2D and 2E), which was that the modulation did not change the magnitude of the line segmentation. This constraint helped us tease out the effect at around 0.24 Hz from around 0.64 Hz – the repetition effect was found at  $\sim 0.24$  Hz, but not at  $\sim 0.64$  Hz (Figure 2E). Therefore, the middle point of the modulation line (3000 ms) had a fixed value equal to half of the amplitude of the sinusoid wave of the line segmentation.

Model 3 was constructed based on Model 1 and included two components: the line segmentation and the onset/offset responses. The peak latencies of the onset/offset responses were fixed at 250 ms and 6300 ms, respectively, which were determined from Figure 3A. The amplitude of the onset response was set to 2 and the offset response was set to 3. The initial values were determined from the observation on Figure 3A where the onset response was roughly twice larger than the neural responses around the first line ending and the offset response was larger than both the onset response and the neural responses around the first line ending. We later varied the amplitude by multiplying a gain with the onset/offset responses and changed the window length – the temporal span of the onset/offset responses. Therefore, there are two free parameters for the onset/offset responses: the window length and the amplitude gain.

Model 4 was a full model and included all the four components. There are four parameters in the full model: line segmentation amplitude (the initial value was set to 1), modulation slope (1), window length of onset/offset responses (1.2 s), and amplitude gain of onset/offset responses (1). The initial values were selected based on our observation that such values enabled the model to generate temporal dynamics similar to the empirical data in Figure 3A (specifically, the first presentation in the left hemisphere). We examined how each parameter affected the effect at  $\sim 0.24$  Hz in the following analyses. It can be seen from the simulation that the full model (Model 4) can best capture the spectral findings in Figure 2D (Figure S2D). The spectral component of the model between 0.12 Hz and 0.28 Hz matched the shape of the amplitude spectra in Figure 2E.

We varied one of the four parameters of the full model while keeping the other three parameters fixed (MATLAB script: *Simulation\_2.m*), so that it showed how each parameter affected the spectral results. We then provided quantitative evaluations on how each parameter affected the effect around 0.24 Hz in the full model (MATLAB script: *Simulation\_3.m*). In Figure 2D, we found the significant repetition effect between 0.12 Hz and 0.28 Hz, so the following analyses were focused on this frequency range. The rationale is that, if we could replicate this repetition effect using our model, we could demonstrate which parameter played a prominent role in modulating the neural component around  $\sim 0.24$  Hz and hence reveal the origin of the repetition effect at  $\sim 0.24$  Hz. Therefore, we averaged the spectral amplitude between 0.12 Hz and 0.28 Hz of the model outputs and tested how each parameter modulated the averaged magnitude within this frequency range.

We first selected a representative value range for each parameter. The selection was motivated by two observations: 1) By comparing the temporal dynamics of the model with the data in Figure 3A, we selected a set of values that produced temporal dynamics similar to the data. For example, if the window length of the onset/offset responses is larger than 1.5 s, the onset/offset responses will overlap with the line segmentation, which is inconsistent with the data. Therefore, we set the upper bound of the window length as 1.5 s; 2) After testing our model using different parameter values, we found that the averaged spectral magnitude approximately had a linear relationship with the parameter values. Therefore, the parameter value ranges would not matter too much to decide how the averaged spectral magnitude changed with the parameter values. Therefore, we selected the value range for each parameter as the following – line segmentation amplitude: 0.1 – 1; modulation slope: 0.1 – 1; window length: 0.15 s – 1.5 s; amplitude gain: 0.1 – 1.

We sampled each parameter in 20 steps within its selected value range and calculated the averaged spectral magnitude corresponding to each selection of the four parameters. Totally, 160000 ( $20 \times 20 \times 20 \times 20$ ) results were generated. We then marginalized three of the four parameters by averaging the results over all the value selections of the three parameters, so we could plot the relationship between the change of one remaining parameter and the change of the spectral magnitude. By marginalizing the parameters this way, we took account into the other three parameters while evaluating the remaining one parameter. We fit a line between each parameter and the spectral magnitude and derived the slopes. We found that the line segmentation amplitude and the modulation slope greatly changed the spectral magnitude between 0.12 Hz and 0.28 Hz, whereas varying the window length and the amplitude gain of onset/offset responses did not change the spectral magnitude to a large extent. It is worth noting that the line segmentation amplitude modulated the magnitude both at  $\sim 0.24$  Hz and at  $\sim 0.64$  Hz. Changing the line segmentation amplitude violated the constraint derived from our data (Figure 2E) that the spectral magnitude of the line segmentation did not change between the two

presentations of the poem stimuli. Therefore, one conclusion can be drawn from the above simulations, which is that the amplitude modulation on the line segmentation from the first to the third line contributed prominently to the effect around 0.24 Hz. Next, we fit our model to the data and explained the repetition effect around 0.24 Hz by manipulating the modulation slope.

Here, we focused on the modulation slope while fixing the other three parameters at their initial values. We fit a line to the data of each condition in Figure 3A from 750 ms to 5250 ms and derived slopes from the data. We took the absolute values of the slopes so that the strength of amplitude modulation indicated by the slopes can be easily compared between conditions. Similar to the results in Figure 2E, we found a repetition effect of slope, with the line segmentation in the first presentation modulated to a larger extent than the second presentation. As we did not find a main effect of hemisphere around 0.24 Hz in Figure 2E, we averaged the slopes over the two hemispheres and simplified our following analyses. We calculated the ratio of the slopes between the first and the second presentations and received a ratio of 3.92.

We applied this ratio from the data to our model and created two conditions corresponding to the two presentations of the poem stimuli. The simulated first presentation had a slope of 1 and the simulated second presentation had a slope of 1/3.92. We calculated the amplitude spectra of the two model presentations and averaged the spectral magnitude between 0.12 Hz and 0.28 Hz, and then calculated the ratio of the model spectral magnitude between the first and the second presentations as well as the ratio from the data for each participant. From the data, we calculated the standard error of the mean of the ratio and derived the 95% confidence interval. Indeed, the ratio of the model spectral magnitude between the two simulated conditions fell within the 95% confidence interval of the data ratio. This demonstrated that the origin of the repetition effect at  $\sim 0.24$  Hz came from the differences of the amplitude modulation on the line segmentation between the first and the second presentations of the poem stimuli.

One remaining issue is whether our model could reproduce the temporal dynamics at  $\sim 0.24$  Hz (Figure 3B, upper left panel). If the answer is positive, this would further validate the explanation provided by the model for the finding at  $\sim 0.24$  Hz. We selected the full model (Model 4) without further changing the initial values of the parameters and employed the same low-pass filter with a cutoff frequency of 0.4 Hz used in our data analyses (*MEG Temporal analyses*) (MATLAB script: *Simulation\_4.m*). Although the detailed dynamics between the filtered model output and the data differed, we replicated in our model the main temporal signature – a dip around the onset of the poem stimuli, a peak around the ending of the first line, and another dip around the ending of the third line at  $\sim 4.5$  s. This result demonstrated that the spectral component of  $\sim 0.24$  Hz emerged from a combination of the onset response to the poem stimuli and the amplitude modulation on the line segmentation.

The simulations conducted above showed that the spectral component of  $\sim 0.24$  Hz was associated with the amplitude modulation on the line segmentation at  $\sim 0.64$  Hz. This modulation effect was the origin of the repetition effect at  $\sim 0.24$  Hz (Figure 2E, upper panel). Our simulation results suggest that the neural amplitude modulation existed across the first three lines. Admittedly, the simulations may not demonstrate this point in a straightforward way. However, the dip around 4.5 s and the amplitude modulation across the first three lines both coincided with the conceptual structure of *Jueju* – ‘Qi (initialization) Cheng (elaboration) Zhuan (continuum/twist) He (integration)’, which is a conceptual grouping between the first three lines and the last line.

## DATA AND CODE AVAILABILITY

The preprocessed MEG dataset, analysis and simulation codes have been deposited in OSF (<https://osf.io/jg3xh/>).

## ADDITIONAL RESOURCES

The poem texts and auditory stimuli, online crowdsourcing behavioral evaluation on the artificial poems, and the behavioral online rating datasets and analysis code can be found at [https://ray306.github.io/neural\\_poem/](https://ray306.github.io/neural_poem/).