Identifying the Periodicity of Information in Natural Language

Anonymous ACL submission

Abstract

Recent theoretical advancement of information density in natural language has brought the following question on desk: To what degree does natural language exhibit *periodicity* pattern in its encoded information? We address this question by introducing a new method called `AutoPeriod` of Surprisal (`APS`). `APS` adopts a canonical periodicity detection algorithm and is able to identify any significant periods that exist in the surprisal sequence of a single document. By applying the algorithm to a set of corpora, we have obtained the following interesting results: Firstly, a considerable proportion of human language demonstrates a strong pattern of periodicity in information; Secondly, new periods that are outside the distributions of typical structural units in text (e.g., sentence boundaries, elementary discourse units, etc.) are found and further confirmed via harmonic regression modeling. We conclude that the periodicity of information in language is a joint outcome from both structured factors and other driving factors that take effect at longer distances. The advantages of our periodicity detection method and its potentials in LLM-generation detection are further discussed.

1 Introduction

Understanding how information transmits in language is a long-standing research topic in linguistics (Shannon, 1948, 1949, 1951; Bell et al., 2003). One of the most influential assumptions is uniform information density (UID) (Aylett and Turk, 2004; Jaeger, 2010; Jaeger and Levy, 2006; Meister et al., 2021), which states that information in language, measured by surprisal, distribute uniformly, as driven by the need for efficient communication, as evidenced in its effect on, e.g., word duration in conversation and word abbreviation (Demberg et al., 2012; Mahowald et al., 2013). Recent work further supplements this assumption with different empirical findings, such as the fluctuation of information with discourse boundaries (Xu and Reitter, 2016, 2018); the thematic structures in dialogue that influence the change of information from speakers (Maës et al., 2022).

These empirical findings give rise to an emerging research topic: the *periodicity* of information in natural language. Periodicity is a bold hypothesis that goes a step further than the simple "fluctuation" of information, which is supported by recent evidence: the spectral features of text surprisal are distinguishable from white noise (Xu and Reitter, 2017; Yang et al., 2023); spectra of surprisal are useful in machine-generation detection (Xu et al., 2024; Liu et al., 2025); surprisal variation can be modeled by harmonic regression, which confirms periodic patterns across various linguistic structures (Tsipidi et al., 2025).

However, there are still some key questions remaining open about this topic. First of all, how can we find **direct evidence for periodicity** of information in language? The Fourier-analysis-based spectral methods proposed in Yang et al. (2023) and Liu et al. (2025) provide useful insights, but can only serve as indirect evidence of periodicity, because the spectrum in frequency-domain was not transformed back to explicit periods in time-domain. On the other hand, the harmonic regression method proposed by Tsipidi et al. (2025), provides a way to validate periodicity in information, given "candidate" periods selected from certain linguistic structures (e.g., sentences, paragraphs etc.). Therefore, it potentially filters out periods that could not be attributed to any pre-defined linguistic units; nor could it precisely tell whether a piece of text is periodic or not, due to its regression analysis nature. This leads to our second question: can we **detect periodicity at document level** with higher quality and more precisely? That is, given a piece of text, we want to find all periods that can pass certain detection threshold with some clearly defined confidence levels, or otherwise to determine its non-existence.

Because the detection is at document level, such method would allow us to explore whether periodicity correlates with other theoretically interesting factors, such as text genres, authorship, and the distinction between human-written texts and those generated by large language models (LLMs).

In this study we address the above-mentioned questions by proposing a periodicity detection algorithm named `AutoPeriod` of Surprisal (`APS`), which is able to directly identify any existing periods in an input sequence of surprisal at certain confidence levels (see Section 3). We apply this algorithm to common linguistic corpora and have obtained positive evidence for periodicity of information (Section 5). The validity of our results is verified with an alternative method (Section 6). The potential merits of periodicity detection in distinguishing LLM-generations (Section 7).

2 Background

2.1 UID and the Natural Fluctuation of Information

The UID theory describes that a general trend of how the information of linguistic units distributes within the course of communication is to stay relatively uniform or constant. However, token-level uniformity/constancy is an asymptotic property that would only manifest when observation is conducted over a sufficiently long sequence, during which the natural fluctuation of information would inevitably occur. For instance, Genzel and Charniak (2002, 2003) observed the increasing surprisal near the beginning of paragraphs and the sudden drop near the end. It can be imagined that if multiple paragraphs of similar lengths are organized together as a chapter, then an overall regular fluctuation (or even *periodicity*) will obviously appear.

Rooted from the above-illustrated coupling between uniformity and fluctuation, recent works have tried to model the latter more formally. Recurring convergence pattern of surprisal is found within topic boundaries in dialogue (Xu and Reitter, 2016, 2018). It is also found that the uniformity trend is only salient in certain contextual units (Giulianelli et al., 2021) and not always observed (Giulianelli and Fernández, 2021). More recently, surprisal fluctuation over discourse is modeled with hierarchically-structured Bayesian regression (Tsipidi et al., 2024) and harmonic regression (Tsipidi et al., 2025).

2.2 Why Studying Periodicity of Surprisal

Given that surprisal can describe how linguistic information changes dynamically in human language, its potential usage has been explored, such as (i) measuring creative use of language such as metaphor and humor (Bunescu and Uduehi, 2022; Xie et al., 2021); (ii) the evaluation of text generation quality (Kharkwal and Muresan, 2014; Yang et al., 2023; Liu et al., 2025); (iii) and recently the detection of LLM-generated text (Xu et al., 2024).

A further study on the periodicity of surprisal will largely facilitate the above-listed research directions. While UID theory and its variants can make asymptotic predictions about the general trends of information in language, using periodicity as a lens would provide more fine-grained insights into the instantaneous dynamics of information, which better meets the purpose of studying the token-by-token generation process. What's missing in the picture is a method and toolkit that can effectively extract periodicity as a clear feature from text—current regression-based methods (Tsipidi et al., 2025) provide ways to validate but not to comprehensively identify periodicity; Fourier-based methods (Xu et al., 2024) lacks direct interpretation. Therefore, the goal of this study is to fill the gap.

3 Periodicity Detection

The method proposed in this study for detecting periodicity in information is based on the `AutoPeriod` algorithm (Vlachos et al., 2005). Originally designed for detecting periodicity in economics and social time series data, this algorithm demonstrates reliable performance in identifying periodic patterns. We adaptively modify and reimplement this algorithm—tailoring its parameters to the characteristics of surprisal—and release it as a standalone toolkit called `AutoPeriod` of Surprisal (`APS`). The algorithm runs two sub-procedures:

1) GetPeriodHints: Get the periodogram spectrum of input sequence and extract the candidate periods called *hints*, whose power values are above a certain threshold obtained from random permutations.

2) ACFFiltering: Filter the period hints using the circular auto-correlation function (ACF), and only keep those that lie on the hills of the ACF curves, which are considered as the valid periods.

We first introduce the mathematical prerequisites for Fourier transform and periodogram in

Section 3.1, as these foundational concepts are essential for understanding the `AutoPeriod` procedure. Readers already familiar with this knowledge can skip directly to Section 3.2 and Section 3.3, where the two sub-procedures are detailed.

3.1 Prerequisites: Fourier transform and periodogram

Given a piece of text that contains N tokens, we first estimate the surprisal of each token by computing the negative log probability $x_n = -\log p(t_n \mid t_{<n})$, where p is the probability of token t_n predicted by a language model based on its preceding context $t_{<n}$. The resulting sequence of surprisal values $\mathbf{x} := x_0, \ldots, x_{N-1}$ will be the input to the discrete Fourier transform (DFT), defined as:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi \frac{k}{N} n}$$

where X_k is the cross correlation of $\mathbf{x_n}$ and a complex sinusoid, $e^{-i2\pi \frac{k}{N} n} = \cos(2\pi \frac{k}{N} n) - i \sin(2\pi \frac{k}{N} n)$ $(n = 0, \ldots, N-1)$.

The norm $\|X_k\| = \sqrt{re(X_k)^2 + im(X_k)^2}$ measures to what degrees the original input can be reconstructed from the periodic signal at frequency $2\pi \frac{k}{N}$. The raw DFT output $\mathbf{X} := X_0, \ldots, X_{N-1}$ has the property of being *even symmetric*, that is, $\|X_k\| = \|X_{N-k}\|$, $\forall k \in 0, \ldots, N-1$. Therefore, it is sufficient to use the first half of $\mathbf{X_k}$, $k \in 0, \ldots, \lceil \frac{N-1}{2} \rceil$ to encode the full spectral information. This truncated magnitude sequence is also known as **periodogram** [1], a widely-used technique in signal processing (Schuster, 1898; Oppenheim et al., 1999). We denote the peridogram as $\mathcal{P} = \{\|X_k\|\}$ $(k \in 0, \ldots, \lceil \frac{N-1}{2} \rceil)$.

3.2 Step 1: Extract hints from periodogram

The GetPeriodHints procedure is described in Algorithm 1. Within this step, the surprisal sequence is first permuted for m times, and the maximum spectrum powers from these permuted sequences are recorded in *maxPower*. The *percentile*-th largest value in *maxPower* is used as the power threshold $\mathcal{P}_{\text{threshold}}$ to find those frequencies/periods whose powers are above this threshold. Therefore, the periods returned are the most significant ones with a *percentile*/100 confidence level.

Algorithm 1 First step of AutoPeriod: Extract period hints.

```
function getPeriodHints(x, m, percentile)
    N ← x.length                          ▷ length of input sequence
    m ← 100                               ▷ number of permutations
    percentile ← 99
    maxPower ← {}                         ▷ empty set
    periodHints ← {}
    for i = 1 to m do
        x^rand ← Permute(x)
        P^rand ← getPeriodogram(x)
        maxPower.add(max(P^rand))
    end for
    maxPower.sort()                       ▷ ascending
    P_threshold ← maxPower[m · (percentile/100)]
    P ← getPeriodogram(x)
    for k = 1 to P.length do
        if P[k] > P_threshold then
            periodHints.add(N/k)
        end if
    end for
    return periodHints
end function
```

Here we use document #0976 from PTB-WSJ as a complete example of analysis. The texts are snippets of TV listings with highly structured format of content – with weekdays followed by channel names and shot summaries (see Figure 1 left). The result of running GetPeriodHints on this document is shown in Figure 1 (right). It can be seen that when confidence level (CL) decreases from .99 to .50, more period hints are extracted. The most confident (CL = .99) period hint occurs at the frequency peak $f \approx .019$, and the corresponding period $T = \frac{1}{f} \approx 53.1$ (tokens), suggesting that a period of surprisal occurs roughly every 53 tokens. CL = .90 is used throughout this study. Before a hint is validated in the next step, we deem it as quasi-periodic.

3.3 Step 2: Period validation via ACF Filtering

The brief version of ACFFiltering procedure at step 2) is described in Algorithm 2. This involves applying a fine-grained validation to the period *hints* found in step 1), which allows spurious periods and false alarms to be filtered out and the *true* period to be identified following further refinement. This

[1] The periodogram algorithm we introduced here follows the design in Vlachos et al. (2005). In our experiments, we tried variants of periodogram, and selected the Lomb-Scargle periodogram (Lomb, 1976; Scargle et al., 1982) to enhance the performance of `APS` in low-frequency range (i.e., long periods).

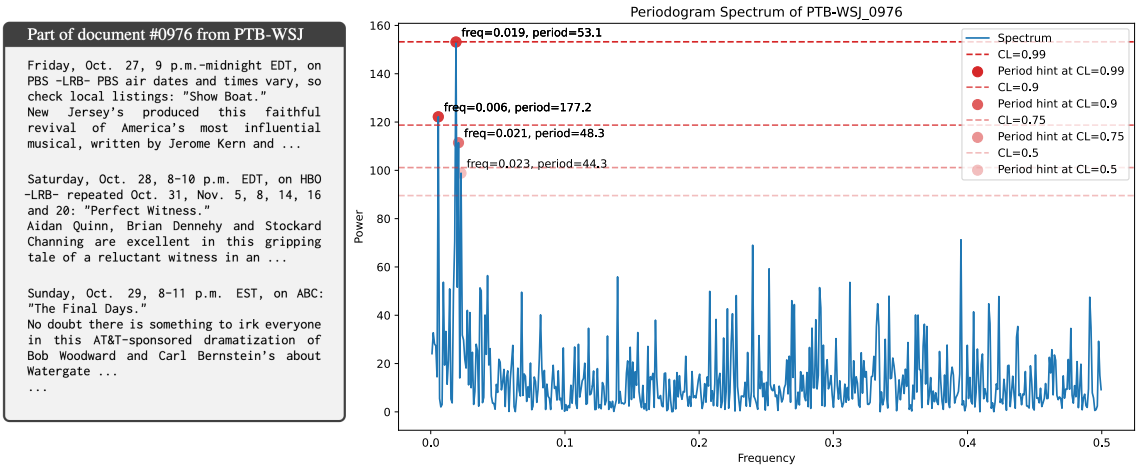

Figure 1: Periodogram spectrum obtained from the surprisal sequence of document #0976 in PTB-WSJ corpus. Right: Red dots indicate the period hints returned by Algorithm 1. Dashed lines are the power thresholds used at different confidence levels.

filtering step is initially designed to increase the detection precision for periodicity in generic time series, and, briefly, the motivation is that a *dominant* period τ of a time series x (of length N) should have a relatively large value of the autocorrelation function (ACF)

$$ACF(\tau) = \frac{1}{N}\sum_{n=0}^{N} x(\tau) \cdot x(n+\tau)$$

which measures how similar the sequence is to its previous values for different τ lags. If a period hint τ from the periodogram lies on a hill of the ACF curve, then we are more confident to consider it as a valid one, otherwise it is likely to be a false alarm. This validation is done to each period hint $\tau = \frac{N}{k}$, by examining the nearby points on the ACF curve within a window of size $W_{N/k} = [a, b]$. If τ does not lie on a hill, this period hint is discarded, otherwise a further *refinement* is done by searching within the nearby region and finding the closest local maximum of the ACF scores. As seen in Figure 2 (bottom), the raw/unrefined period hints $\tau'_1 = 53.2$ and $\tau'_2 = 177.2$ are refined to the final integer periods $\tau_1 = 52$ and $\tau_2 = 163$, respectively.

Algorithm 2 ACF Filtering (brief)

```
function ACFFiltering(x, periodHints)
    N ← x.length
    ACF ← getACF(x)
    M ← periodHints.length
    refinedPeriods ← {}                ▷ empty set
    for i = 1 to M do
        τ ← N/k ← periodHints[i]
        τ' ← local maximum within a range
        if τ' is on hill of ACF then
            refinedPeriods.add(τ')
        end if
    end for
    return refinedPeriods
end function
```

4 Corpora and Language Models

This study utilizes several text corpora. Penn Treebank (PTB) (Marcus et al., 1993) and Chinese Treebank[2] (CTB) (Xue et al., 2013) are adopted as representatives of human-written texts in English and Chinese, respectively. Specifically, the Wall Street Journal (WSJ) and Brown portions are used for PTB. Moreover, to explore the relationship between the global periodicity detected by APS and the local linguistic structures focused on by Tsipidi et al. (2025), we incorporate the RST Discourse Treebank (Carlson et al., 2001) for English and the GCDT corpus (Peng et al., 2022) for Chinese. These corpora provide discourse-level annotations, enabling a comparative analysis. The detection results of the corpora mentioned above are mainly discussed in Section 5 and Section 6.

To investigate the periodicity differences between human-written and LLM-generated texts, we use two datasets: FACE collected by Liu et al. (2025) and EvoBench by Yu et al. (2025). Our analysis focuses on the English news and Chinese news por-

[2] https://catalog.ldc.upenn.edu/LDC2013T21

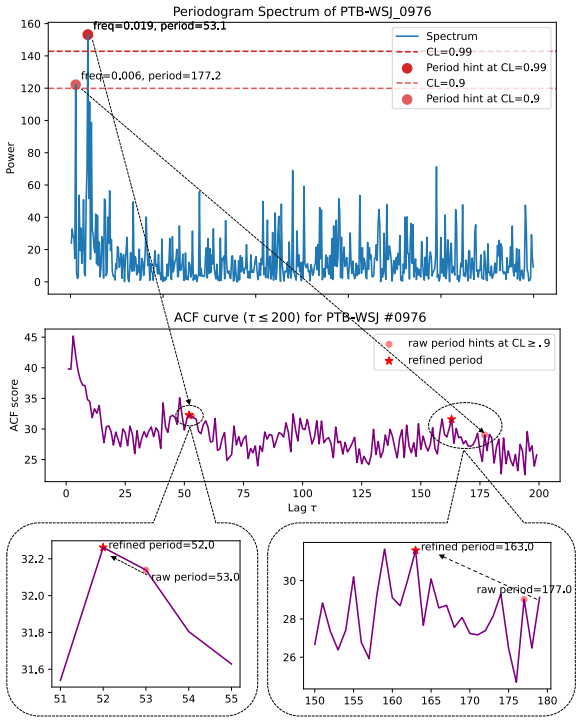

Figure 2: ACF filtering for surprisal sequence of document #0976 in PTB-WSJ corpus. Hints (red dots in top) are mapped to the raw periods in the ACF curve (red dots in middle). ACF filtering refines and returns the final periods (red stars in bottom).

tions of FACE, and the XSum portion of EvoBench. For each portion, we select a subset generated by one of the top-performing LLMs. The specific subsets used are detailed in Table 8, with results presented in Section 7.

To extract surprisal sequences from the corpora, LLMs are used to encode the input texts. For English texts, LLaMA3-8B (Touvron et al., 2023) is adopted. For Chinese texts, Qwen2-7B (Yang et al., 2024) is adopted. Specifically, to ensure comparability with the experiments of Tsipidi et al. (2025), we use Yarn-LLaMA2-7B (Peng et al., 2024) for RST Discourse Treebank.

5 Result: Identified Periods of Information

5.1 Proportions of periodic documents

Applying APS on a text corpus, indicated by Σ, would partition it into three subsets: P_1, the set of documents that contains at least one hint (not necessarily valid); P_2, the set of documents that contain at least one valid period; and lastly, $\Sigma - P_2$, the remaining set containing no hints or periods (see Figure 3).

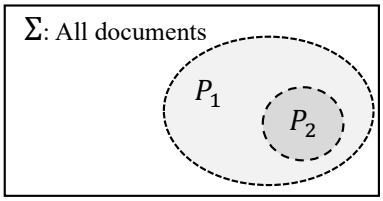

Figure 3: Partitioning a corpus based on the periodicity detection outcomes from APS.

Across the corpora examined, we consistently find that **a non-negligible proportion of texts exhibit strong periodicity of information**. The counts of documents $|\Sigma|$, $|P_1|$ and $|P_2|$, and the proportions $|P_1|/|\Sigma|$, $|P_2|/|P_1|$ and $|P_2|/|\Sigma|$ are shown in Table 1. P_2 are the most strictly periodic documents, whose proportion $|P_2|/|\Sigma|$ reaches above 5% for all corpora. P_1 is the quasi-periodic set, and its odds $|P_1|/|\Sigma|$ is significantly higher, reaching about 15% on average. The relatively high ratio of $|P_2|/|P_1|$ (about 66%) indicates that once a hint is detected by GetPeriodHints (step 1), it has a large chance to pass ACFFiltering (step 2).

All three ratios are higher in Chinese corpora than in English. The exceptionally high values of GCDT against the rest also suggests that text genre is an important factor affecting periodicity – GCDT contains five genres of formal texts, while WSJ, Brown, and CTB are solely collected from news articles. We will leave the comprehensive examination on how text genre (such as the dimensions of verse vs. prose, oral vs. written, etc.) influence the periodicity of information as future work.

5.2 Comparing APS periods with linguistic structural units

To assess the relationship between the global periodicity detected by APS and the local linguistic structures focused on by Tsipidi et al. (2025), we conducted a quantitative comparison between the lengths of periods identified by APS and that of structural units (EDUs, sentences, paragraphs) annotated in the RST Discourse Treebank. The distribution of the detected periods is presented in Figure 4, and the average lengths of structural units are marked along the x-axis. A prominent peak in the period distribution is observed around 25 tokens, which aligns closely with the average lengths of the three structural units. This alignment suggests that a subset of the periodic patterns detected can indeed be attributed to these fundamental structural units.

However, a substantial proportion of the detected

periods fall outside the typical length distributions of the considered structural units. Notably, while the *paragraph* (the longest among the three) averages around 70 tokens, our method identifies a significant number (approximately 50%) of periods exceeding 100 tokens in length. The existence of these long-period patterns cannot be accounted for by the structural units. This clear mismatch indicates that **the global periodicity of information identified by APS is not fully explicable by the succession of local discourse structures alone**. Instead, it suggests the presence of additional, yet-to-be-identified, factors contributing to the long-periodicity of information in text.

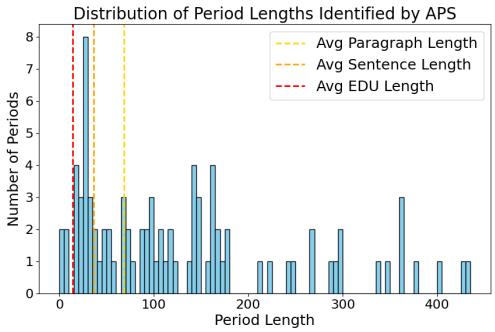

Figure 4: Distribution of periods identified by APS on RST Discourse Treebank, compared with the average lengths of structural units (EDUs, sentences, paragraphs). The average lengths of structural units are marked as dashed lines.

6 Result: Sanity-Check with Harmonic Regression Modeling

In this section, we use an alternative method to carry out a sanity-check on the detection results from APS. Tsipidi et al. (2025) formulates periodicity of information as the effect of harmonic components in predicting the surprisal of a token w_t using harmonic regression (HR) models. In their models, periods are defined as the lengths of structural units (sentences, paragraphs, and EDUs). The prediction task is formulated as:

$$s(w_t) \sim \text{baseline} + \text{HR}(U_t) \tag{1}$$

in which "baseline" indicates the same basic predictors (e.g., relative positions of w_t, etc.) used in Tsipidi et al. (2025), and the $\text{HR}(U_t)$ term is the linear combination of sine and cosine components corresponding to a certain unit length U_t:

$$\text{HR}(U_t) = \beta_0 + \sum_{k=1}^{K} \left(\beta_{1,k} \cdot \sin\left(\frac{k2\pi t}{U_t}\right) + \beta_{2,k} \cdot \cos\left(\frac{k2\pi t}{U_t}\right) \right) \tag{2}$$

Significant coefficients in the form of amplitude $A_k = \sqrt{\beta_{1,k}^2 + \beta_{2,k}^2}$ from the $\text{HR}(U_t)$ terms would indicate that certain degrees of periodicity exist within the scope of U_t. Based on this paradigm, we proceed with two inquires:

1. Does using the hints and periods from APS (our method) as the scaling factor U_t result in significant HR modeling outcomes?

2. Do the documents that return positive in APS detection result in better overall performance in HR modeling?

6.1 Periods/hints directly used as scaler in HR

We replace U_t in Equation (1) with the *hints* and *periods* returned by APS, and then examine the corresponding HR modeling outcomes. If there are multiple hints or periods in one document, then the largest one is selected. The modeling formulas then become:

$$m_{\text{hints}} : s(w_t) \sim \text{baseline} + \text{HR}(\textit{hints})$$
$$m_{\text{periods}} : s(w_t) \sim \text{baseline} + \text{HR}(\textit{periods})$$

The models m_{hints} and m_{period} are fit against the P_1 and P_2 portions of WSJ, respectively (to be directly comparable to Tsipidi et al.'s (2025) results, only WSJ is used). The coefficients of the harmonic terms $\text{HR}(\textit{hints})$ and $\text{HR}(\textit{periods})$ with the top-3 largest amplitude values are reported in Tables 2 and 3. The results show that both terms are indeed significant predictors in HR modeling, which confirms the periodicity of hints and periods returned by APS.

There is no significant differences between the A_k values of hints and periods. This suggests that, although the latter is stronger than the former by the standard of passing the periodicity threshold, their modeling effects on predicting local surprisal are almost equivalent. We also find that the A_k values from hints and periods are smaller than those from EDU, the best predictor reported in Tsipidi et al. (2025). This is to be expected, as hints/periods are fixed values within a document and are therefore less dependent on the position of

Table 1: Counts of document with non-empty period hints ($|P_1|$) and valid periods ($|P_2|$) identified by APS, along with corresponding ratios relative to the total document counts ($|\Sigma|$) across four corpora. Last two rows are the mean and median lengths of the hints/periods identified in P_1/P_2, respectively.

	WSJ_{en}	Brown_{en}	GCDT_{zh}	CTB_{zh}	Avg.
$\lvert\Sigma\rvert$	2499	500	50	2773	-
$\lvert P_1\rvert$	221	52	15	304	-
$\lvert P_2\rvert$	131	25	13	212	-
$\lvert P_1\rvert/\lvert\Sigma\rvert$	8.84%	10.40%	30.00%	10.96%	15.05%
$\lvert P_2\rvert/\lvert P_1\rvert$	59.28%	48.08%	86.67%	69.74%	65.94%
$\lvert P_2\rvert/\lvert\Sigma\rvert$	5.24%	5.00%	26.00%	7.65%	10.97%
P_1: mean (median)	127.68 (100)	111.37 (28)	113.76 (81)	97.53 (51)	112.36 (74)
P_2: mean (median)	135.65 (108)	144.57 (84)	110.53 (56)	89.69 (34)	113.14 (70)

tokens. This naturally makes them less informative predictors. However, this does not affect their periodicity. Rather, it strengthens our conclusion that APS can detect global periodic patterns that differ from local units.

Table 2: Significance test of the top-3 harmonic components with the largest amplitudes in HR models using APS *periods* as scaling factors on WSJ (P_2 portion).

k	A_k	β	coef	$P > \lvert t\rvert$
1	0.2022	$\beta_{1,1}$	0.1213	0.000
		$\beta_{2,1}$	0.1571	0.000
3	0.0652	$\beta_{1,3}$	0.0524	0.000
		$\beta_{2,3}$	0.0405	0.000
4	0.0467	$\beta_{1,4}$	0.0367	0.002
		$\beta_{2,4}$	0.0289	0.010

Table 3: Significance test of the top-3 harmonic components with the largest amplitudes in HR models using APS *hints* as scaling factors on WSJ (P_1 portion).

k	A_k	β	coef	$P > \lvert t\rvert$
1	0.1772	$\beta_{1,1}$	0.0736	0.000
		$\beta_{2,1}$	0.1612	0.000
5	0.0731	$\beta_{1,5}$	0.0503	0.000
		$\beta_{2,5}$	-0.0531	0.000
7	0.0575	$\beta_{1,7}$	0.0036	0.723
		$\beta_{2,7}$	-0.0574	0.000

6.2 Filtering data using periods/hints improves HR performance

To test whether APS can filter out data with better overall periodicity, we fit HR models on four different portions, $\{P_2, P_1, \Sigma, \Sigma - P_1\}$ of RST and GCDT corpora, using three structural units (EDU, sentence, and document) as different scaling factors. We assume that the resulting mean squared errors (MSE) should rank in a descending order, i.e., $P_2 < P_1 < \Sigma < \Sigma - P_1$, as P_2 is the most periodic portion according to APS, while $\Sigma - P_1$ is the least.

The results in Tables 4 and 5 confirm our assumption. The rankings remain consistent across different scaling factors and even so in baseline models (without HR terms). This indicates that the documents identified by APS as containing strong periodicity of information are more predictable in terms of token surprisal. Scaling by structural units further improves the performance on the filtered portions, and EDU scaling is the most effective one, which is consistent with Tsipidi et al.'s (2025) results. From this perspective, the global periodicity identified by APS and the local periodicity modeled by HR using structural units are compatible. It is possible that the periodic patterns of discourse structures contribute to the overall periodicity of information in text.

7 Result: Human vs. Generated Texts

By applying APS on the corpora of human-written and LLM-generated text pairs, we can observe periodicity differences between them on a dataset-level scale. These data are gathered by obtaining a generated response from LLMs by feeding the beginning of a human-written text as a prompt. In

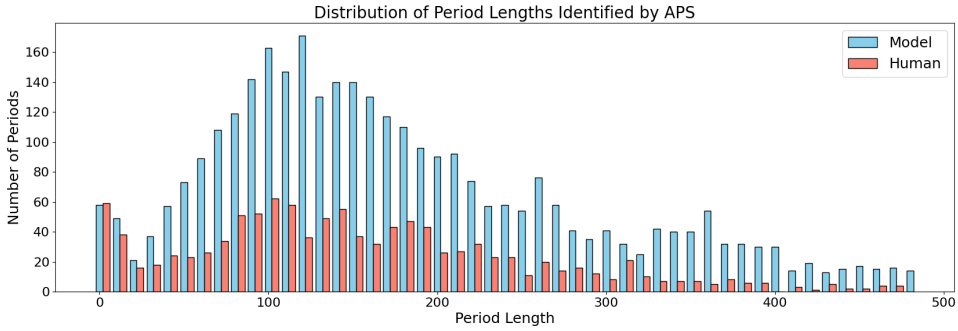

Figure 5: Distributions of periods from human (FACE BBC News) and LLM-generated (LLaMA-3-70B) texts.

Table 4: MSE of HR modeling on WSJ using different scaling factors and different data portions.

	Baseline	EDU	Sent.	Doc.
P_2	16.3273	**13.7788**	15.5958	16.2915
P_1	16.8654	**14.1434**	16.1400	16.8353
Σ	18.2635	15.1843	17.5619	18.2470
$\Sigma - P_1$	18.5859	15.4160	17.8875	18.5692

Table 5: MSE of HR modeling on GCDT using different scaling factors and different data portions.

	Baseline	EDU	Sent.	Doc.
P_2	6.8699	**6.3902**	6.8070	6.8489
P_1	6.9967	**6.5097**	6.9303	6.9789
Σ	7.2485	6.5961	7.1511	7.2279
$\Sigma - P_1$	7.2832	6.5644	7.1673	7.2554

general, we find that LLM-generated texts contain stronger periodicity of information than human-written ones in both English and Chinese. For instance, the proportion of documents with valid periods in LLaMA3-70B-generated fake BBC news is 30.06%, which is more than twice that of the real news (14.80%). Detailed results are listed in Table 8.

Figure 5 shows the distributions of the detected periods in human-written and LLM-generated texts in FACE BBC News data. It can be observed that the LLM-generated texts exhibit more periodic patterns than human-written texts, especially in the long-period range (> 50 tokens). A possible reason could be that LLMs are more capable of keeping the global coherence when generating within long ranges. Human writers, however, are limited by the memory capacity, and hence are less likely to maintain consistency under large spans.

The above finding lends interpretation to some existing studies that utilize periodicity of information as a feature for the task of detecting model-generated texts (Xu et al., 2024) and for automatic NLG evaluation (Liu et al., 2025). Through APS, we can explicitly tell *how* the periodicity of information in LLM-generated texts is different from that in human-written ones. Therefore, information periodicity holds promise as a distinctive feature for generation-related tasks.

8 Conclusions

In this study, we propose AutoPeriod of Surprisal (APS), an algorithm for revealing periods of surprisal in natural language with a high resolution, which contributes to the emerging research topic of periodicity of information in human language. Our main observations and conclusions are as follows:

Periodicity can be detected. Our method confirms that albeit a subtle phenomenon, the periodicity of information can be detected at the single document level. It is validated that there is a considerable proportion of human written language that exhibit significant periodicity of information.

Sources of periodicity. We find evidence for periods that are not purely due to the linguistic structures, such as sentences, paragraphs, and EDUs. It indicates there are other factors than structural information processing pressure that cause the periodicity, which needs be studied in future work.

Usability in human-machine distinction. APS is not a regression-based technique that focuses on analysis, but rather a detection method that attempts to extract useful cues about how information distributes temporally given any input text. Therefore, it has potentials in distinguishing human- vs. model-generated texts, under the basic assumption that human language production is a cognitively-constrained process.

Limitations

Within the scope of this study, we consider the following limitations, which we believe can be further addressed in future work. Firstly, this study only examines English and Chinese texts. We believe that different languages may have different language structure, as the consequence we have considered, our proposed `APS` method might have varied performance across different texts from different language.

Secondly, our detection results in Section 5.1 suggest that the periodicity of information is influenced by text genre. We also find that parts of the detected periods cannot be explained by local structural units, as discussed in Section 5.2. However, we do not further analyze the sources of these periods or those potential factors that influence the periodicity of information. We leave these inquiries to future work.

Thirdly, the performance of `AutoPeriod` might not be as accurate in detecting signals with short periods as those with long periods. The corresponding frequencies of the component signals returned by DFT distribute unevenly on the axis of period length. Components with short periods are more densely distributed than those with long periods. Therefore, when searching for period hints by random perturbation, short periods are more likely to be false positives. From the distribution of the detected periods, we observe that the prominent peak in the region of short periods tends to shrink when the confidence level is set higher. The observation indicates that the detection of short periods is less robust.

References

Matthew Aylett and Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1):31–56.

Alan Bell, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory, and Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, 113(2):1001–1024.

Razvan C. Bunescu and Oseremen O. Uduehi. 2022. Distribution-based measures of surprise for creative language: Experiments with humor and metaphor. In *Proceedings of the 3rd Workshop on Figurative Language Processing (FLP)*, pages 68–78, Abu Dhabi, UAE (Hybrid). Association for Computational Linguistics.

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. 2001. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, page 1–10, Aalborg, Denmark. Association for Computational Linguistics.

Vera Demberg, Asad Sayeed, Philip Gorinski, and Nikolaos Engonopoulos. 2012. Syntactic surprisal affects spoken word duration in conversational contexts. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 356–367, Jeju Island, Korea. Association for Computational Linguistics.

Dmitriy Genzel and Eugene Charniak. 2002. Entropy rate constancy in text. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 199–206, Philadelphia, PA, USA. Association for Computational Linguistics.

Dmitriy Genzel and Eugene Charniak. 2003. Variation of entropy and parse trees of sentences as a function of the sentence number. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pages 65–72.

Mario Giulianelli and Raquel Fernández. 2021. Analysing human strategies of information transmission as a function of discourse context. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 647–660, Online. Association for Computational Linguistics.

Mario Giulianelli, Arabella Sinclair, and Raquel Fernández. 2021. Is information density uniform in task-oriented dialogues? In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 8271–8283, Punta Cana, Dominican Republic. Association for Computational Linguistics.

T. Florian Jaeger. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1):23–62.

T. Florian Jaeger and Roger Levy. 2006. Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems*, 19.

Gaurav Kharkwal and Smaranda Muresan. 2014. Surprisal as a predictor of essay quality. In *Proceedings of the 9th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 54–60.

Zhichen Liu, Yongyuan Li, Yang Xu, Yu Wang, Yingfang Yuan, and Zuhao Yang. 2025. Evaluating open-ended text generation of large language models using spectral distances of surprisal. In *Findings of the Association for Computational Linguistics: EMNLP 2025*.

N R Lomb. 1976. Least-squares frequency analysis of unequally spaced data.

Eliot Maës, Philippe Blache, and Leonor Becerra. 2022. Shared knowledge in natural conversations: can entropy metrics shed light on information transfers? In *Proceedings of the 26th Conference on Computational Natural Language Learning (CoNLL)*, pages 213–227, Abu Dhabi, UAE (Hybrid). Association for Computational Linguistics.

Kyle Mahowald, Evelina Fedorenko, Steven T. Piantadosi, and Edward Gibson. 2013. Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, 126(2):313–318.

Mitch Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn treebank. *Computational Linguistics*, 19(2):313–330.

Clara Meister, Tiago Pimentel, Patrick Haller, Lena Jäger, Ryan Cotterell, and Roger Levy. 2021. Revisiting the Uniform Information Density hypothesis. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 963–980, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Alan V. Oppenheim, Ronald W. Schafer, and John R. Buck. 1999. *Discrete-Time Signal Processing*. Prentice Hall, Upper Saddle River, NJ, USA.

Bowen Peng, Jeffrey Quesnelle, Honglu Fan, and Enrico Shippole. 2024. YaRN: Efficient context window extension of large language models. In *Proceedings of the 12th International Conference on Learning Representations*.

Siyao Peng, Yang Janet Liu, and Amir Zeldes. 2022. GCDT: A Chinese RST treebank for multigenre and multilingual discourse parsing. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing*, pages 382–391, Online. Association for Computational Linguistics.

Scargle, J., and D. 1982. Studies in astronomical time series analysis. ii - statistical aspects of spectral analysis of unevenly spaced data. *ApJ*, 263(2):835–835.

Arthur Schuster. 1898. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terrestrial Magnetism*, 3(1):13–41.

Claude E. Shannon. 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.

Claude E. Shannon. 1949. Communication theory of secrecy systems. *The Bell System Technical Journal*, 28(4):656–715.

Claude E. Shannon. 1951. The redundancy of English. In *Transactions of the 7th Conference on Cybernetics*, pages 248–272, New York, NY, USA.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. LLaMA: Open and efficient foundation language models. *Preprint*, arXiv:2302.13971.

Eleftheria Tsipidi, Samuel Kiegeland, Franz Nowak, Tianyang Xu, Ethan Wilcox, Alex Warstadt, Ryan Cotterell, and Mario Giulianelli. 2025. The harmonic structure of information contours. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 31636–31659, Vienna, Austria. Association for Computational Linguistics.

Eleftheria Tsipidi, Franz Nowak, Ryan Cotterell, Ethan Wilcox, Mario Giulianelli, and Alex Warstadt. 2024. Surprise! Uniform Information Density isn't the whole story: Predicting surprisal contours in long-form discourse. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 18820–18836, Miami, Florida, USA. Association for Computational Linguistics.

Michail Vlachos, Philip S. Yu, and Vittorio Castelli. 2005. On periodicity detection and structural periodic similarity. In *Proceedings of the 2005 SIAM International Conference on Data Mining*, pages 449–460, Newport Beach, CA, USA. SIAM.

Yubo Xie, Junze Li, and Pearl Pu. 2021. Uncertainty and surprisal jointly deliver the punchline: Exploiting incongruity-based features for humor recognition. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 33–39, Online. Association for Computational Linguistics.

Yang Xu and David Reitter. 2016. Entropy converges between dialogue participants: Explanations from an information-theoretic perspective. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 537–546, Berlin, Germany. Association for Computational Linguistics.

Yang Xu and David Reitter. 2017. Spectral analysis of information density in dialogue predicts collaborative task performance. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 623–633, Vancouver, Canada. Association for Computational Linguistics.

Yang Xu and David Reitter. 2018. Information density converges in dialogue: Towards an information-theoretic model. *Cognition*, 170:147–163.

Yang Xu, Yu Wang, Hao An, Zhichen Liu, and Yongyuan Li. 2024. Detecting subtle differences between human and model languages using spectrum of relative likelihood. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 10108–10121, Miami, Florida, USA. Association for Computational Linguistics.

Nianwen Xue, Xiuhong Zhang, Zixin Jiang, Martha Palmer, Fei Xia, Fu-Dong Chiou, and Meiyu Chang. 2013. Chinese treebank 8.0 ldc2013t21. *Linguistic Data Consortium, Philadelphia*.

An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, and 1 others. 2024. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*.

Zuhao Yang, Yingfang Yuan, Yang Xu, Shuo Zhan, Huajun Bai, and Kefan Chen. 2023. FACE: Evaluating natural language generation with Fourier analysis of cross-entropy. In *Advances in Neural Information Processing Systems*, volume 36, pages 17038–17056. Curran.

Xiao Yu, Yi Yu, Dongrui Liu, Kejiang Chen, Weiming Zhang, Nenghai Yu, and Jing Shao. 2025. EvoBench: Towards real-world LLM-generated text detection benchmarking for evolving large language models. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 14605–14620, Vienna, Austria. Association for Computational Linguistics.

A Implementation details of APS

A full version of the APS algorithm is presented in Algorithm 3. Compared with the pseudo-code in Algorithm 2, it details the process of determining the search window W, the linear regression based hill validation, and the final period selection strategy.

B Amplitudes and corresponding p-values of HR modeling

To verify the periodicity of information identified by APS, we fit HR models using the hints and periods returned by APS as the scaling factors, and examine the significance of the harmonic terms. We simplify the experiment setting of HR modeling, picking 10 terms ($K = 10$) in Equation (2), and fitting the model on the entire document set of WSJ. The amplitudes A_k and the corresponding p-values of the coefficients $\beta_{1,k}$ and $\beta_{2,k}$ are reported in Tables 6 and 7.

Table 6: Significance test of all the amplitudes in HR models using APS *hints* as scaling factors on WSJ (P_1 portion).

k	A_k	β	coef	$P > \|t\|$
1	0.1772	$\beta_{1,1}$	0.0736	0.000
		$\beta_{2,1}$	0.1612	0.000
2	0.0255	$\beta_{1,2}$	0.0228	0.023
		$\beta_{2,2}$	0.0115	0.227
3	0.0539	$\beta_{1,3}$	0.0335	0.001
		$\beta_{2,3}$	0.0423	0.000
4	0.0462	$\beta_{1,4}$	0.0246	0.016
		$\beta_{2,4}$	0.0391	0.000
5	0.0731	$\beta_{1,5}$	0.0503	0.000
		$\beta_{2,5}$	-0.0531	0.000
6	0.0373	$\beta_{1,6}$	0.0034	0.749
		$\beta_{2,6}$	0.0371	0.000
7	0.0575	$\beta_{1,7}$	0.0036	0.723
		$\beta_{2,7}$	-0.0574	0.000
8	0.0318	$\beta_{1,8}$	0.0304	0.003
		$\beta_{2,8}$	0.0093	0.326
9	0.0191	$\beta_{1,9}$	0.0163	0.113
		$\beta_{2,9}$	0.0099	0.290
10	0.0406	$\beta_{1,10}$	0.0343	0.001
		$\beta_{2,10}$	-0.0217	0.021

C Full results of human vs. generated texts

We provide the full results of the detection of periods in human-written vs. LLM-generated texts in Table 8, including the counts and ratios of documents with non-empty period hints ($|P_1|$) and valid periods ($|P_2|$) identified by APS in three corpora. For EvoBench, we choose the XSum portion of the dataset, and select the subset generated by GPT-4o as the LLM-generated texts. For FACE, we choose the News portion of the dataset in both English and Chinese, and select the texts generated by LLaMA3-70B and Qwen2-72B as the LLM-generated texts, respectively.

Table 7: Significance test of all the amplitudes in HR models using APS *periods* as scaling factors on WSJ (P_2 portion).

k	A_k	β	coef	$P > \|t\|$
1	0.2022	$\beta_{1,1}$	0.1273	0.000
		$\beta_{2,1}$	0.1571	0.000
2	0.0449	$\beta_{1,2}$	0.0316	0.007
		$\beta_{2,2}$	0.0319	0.005
3	0.0662	$\beta_{1,3}$	0.0524	0.000
		$\beta_{2,3}$	0.0405	0.000
4	0.0467	$\beta_{1,4}$	0.0367	0.002
		$\beta_{2,4}$	0.0289	0.010
5	0.0466	$\beta_{1,5}$	0.0453	0.000
		$\beta_{2,5}$	-0.0111	0.318
6	0.0296	$\beta_{1,6}$	0.0282	0.018
		$\beta_{2,6}$	-0.0090	0.420
7	0.0216	$\beta_{1,7}$	0.0190	0.108
		$\beta_{2,7}$	-0.0101	0.369
8	0.0299	$\beta_{1,8}$	0.0133	0.262
		$\beta_{2,8}$	-0.0268	0.018
9	0.0310	$\beta_{1,9}$	0.0234	0.050
		$\beta_{2,9}$	0.0203	0.069
10	0.0083	$\beta_{1,10}$	0.0036	0.762
		$\beta_{2,10}$	-0.0075	0.504

Algorithm 3 ACF Filtering

```
function ACFFiltering(x, periodHints)
    N ← x.length
    ACF ← getACF(x)
    M ← periodHints.length
    refinedPeriods ← {}
    for i = 1 to M do
        τ ← periodHints[i]
        τ ← N/k ← periodHints[i]
                                    ▷ Determine search window using adjacent periods
        τ_prev ← N/(k − 1)
        τ_next ← N/(k + 1)
        W ← [(τ+τ_next)/2 − 1, (τ+τ_prev)/2 + 1]
                                    ▷ Find optimal split point using linear regression
        ε_min ← ∞
        t_best ← −1
        slope_L ← 0, slope_R ← 0
        for each t in W do
            Split W at t into L and R
            slope_L^t, ε_L^t ← LinearRegression(L, ACF[L])
            slope_R^t, ε_R^t ← LinearRegression(R, ACF[R])
            if ε_L^t + ε_R^t < ε_min then
                ε_min ← ε_L^t + ε_R^t
                t_best ← t
                slope_L ← slope_L^t, slope_R ← slope_R^t
            end if
        end for
                                    ▷ Validate if the split forms a hill
        θ_L ← arctan(slope_L)/(π/2)
        θ_R ← arctan(slope_R)/(π/2)
        Δθ ← |θ_L − θ_R|
        isHill ← (slope_L > slope_R) ∧ (Δθ > 0.01)
        if isHill then
            refinedPeriods.add(t_best)
        end if
    end for
    return refinedPeriods
end function
```

Table 8: Counts and ratios of documents with non-empty period hints ($|P_1|$) and valid periods ($|P_2|$) identified by APS in human-written vs. LLM-generated texts across three corpora.

	$\text{EvoBench}_{\text{en}}$		FACE_{en}		FACE_{zh}	
	Human	GPT-4o	Human	LLaMA3-70B	Human	Qwen2-72B
$\lvert\Sigma\rvert$	150	150	5000	5000	5000	5000
$\lvert P_1\rvert$	9	19	1032	1882	1209	1592
$\lvert P_2\rvert$	5	11	740	1503	943	1231
$\lvert P_1\rvert/\lvert\Sigma\rvert$	6.00%	12.67%	20.64%	37.64%	24.18%	31.84%
$\lvert P_2\rvert/\lvert P_1\rvert$	55.56%	57.89%	71.71%	79.86%	78.00%	77.32%
$\lvert P_2\rvert/\lvert\Sigma\rvert$	3.33%	7.33%	14.80%	30.06%	18.86%	24.62%