Theoretical Computer Science 550 (2014) 66–78

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Palindromes in circular words

Jamie Simpson [a,b,*]

[a] Department of Mathematics and Statistics, Murdoch University, 90 South Street, Murdoch, Western Australia 6150, Australia
[b] Department of Mathematics and Statistics, Curtin University of Technology, GPO Box U1987, Perth, Western Australia 6845, Australia

ARTICLE INFO

Article history:
Received 16 December 2013
Received in revised form 3 July 2014
Accepted 9 July 2014
Available online 22 July 2014
Communicated by M. Crochemore

Keywords:
Circular words
Palindromes

ABSTRACT

There is a very short and beautiful proof that the number of distinct non-empty palindromes in a word of length n is at most n. In this paper we show, with a very complicated proof, that the number of distinct non-empty palindromes with length at most n in a circular word of length n is less than $5n/3$. For n divisible by 3 we present circular words of length n containing $5n/3-2$ distinct palindromes, so the bound is almost sharp. The paper finishes with some open problems.

1. Introduction

We use the usual notation for combinatorics on words. A word of n letters is $x = x[1..n]$, with $x[i]$ being the ith letter and $x[i..j]$ the *factor* of elements from position i to position j. If $i = 1$ then the factor is a *prefix* and if $j = n$ it is a *suffix*. A factor which is neither a suffix nor a prefix is *proper*. If $i_1 \le i_2 \le j_1 \le j_2$ then the *union* of $x[i_1..j_1]$ and $x[i_2..j_2]$ is $x[i_1..j_2]$. The letters in x come from some *alphabet* A. The *length* of x, written $|x|$, is the number of letters that x contains. If $w = uv$ then vu is a *conjugate* of w. The *empty word* ϵ is a word with length 0. A word x or factor x is *periodic* with *period* p if $x[i] = x[i+p]$ for all i such that $x[i]$ and $x[i+p]$ are in x. We will use the following well-known propositions.

Lemma 1. *(See [6, Periodicity Lemma].) Let w be a word having two periods p and q. If $|w| \ge p+q-\gcd(p,q)$ then w also has period $\gcd(p,q)$.*

Lemma 2. *Let w be a word having two periods p and q with $q < p$. Then the suffix and prefix of length $|w|-q$ both have period $p-q$.*

This is [9, Lemma 8.1.1], [8, Lemma 2.1] and is extended in [7, Lemma 2].

Lemma 3. *(See [9, Lemma 8.1.3].) Let w be a word with period q which has a factor u with $|u| \ge q$ that has period r, where r divides q. Then w has period r.*

Lemma 4. *(See [9, Lemma 8.1.2].) If u, v and w are words such that uv and vw both have period p and $|v| \ge p$ then the word uvw has period p.*

* Correspondence to: Department of Mathematics and Statistics, Curtin University of Technology, GPO Box U1987, Perth, Western Australia 6845, Australia.
E-mail address: Jamie.Simpson@curtin.edu.au.

http://dx.doi.org/10.1016/j.tcs.2014.07.012
0304-3975/

The *reversal* of a word $w[1\,..\,n]$ is the word $\overline{w} = w[n]w[n-1]\,..\,w[1]$, and w is a *palindrome* if $w = \overline{w}$. The empty word ϵ is a palindrome, however in this paper we will only be concerned with non-empty palindromes. A palindrome is *odd* or *even* if its length is, respectively, odd or even. If $w[i\,..\,j]$ is a palindrome we say it has *centre* $C = (i+j)/2$ and *radius* $R = (j-i)/2$. Note that C and R are integers if the palindrome is odd and each is an integer plus 1/2 if the palindrome is even. In either case $2C$, $2R$ and $C+R$ are integers. We write $P(C,R)$ for the palindrome with centre C and radius R. Thus if $P(C,R)$ is a palindrome in a word w then

$$P(C,R) = w[C-R\,..\,C+R]$$

and the palindrome has length $2R+1$. We will sometimes use $P(C,R)$ to refer simply to the interval $[C-R, C+R]$. It will be clear from the context when this is so. For $i \in P(C,R)$ we have

$$w[i] = w[2C-i] \tag{1}$$

and

$$\overline{w\big[\lfloor C\rfloor\,..\,C+R\big]} = w\big[C-R\,..\,\lceil C\rceil\big]. \tag{2}$$

If $w[i\,..\,j]$ is a palindrome with $j \ge i+2$ then so is $w[i+1\,..\,j-1]$. If $w[C-R\,..\,C+R]$ is a palindrome but none of $w[C-R-1\,..\,C+R+1]$, $w[C-R\,..\,C+R+1]$, or $w[C-R-1\,..\,C+R]$ is, then we say $w[C-R\,..\,C+R]$ is a *maximal palindrome*. The second and third cases here mean that we do not, for example, consider *aa* to be maximal in *baaac*. If $w[C-R\,..\,C+R]$ is maximal then the palindromes $w[C-R+i\,..\,C+R-i]$, $i = 1,\dots,\lfloor R\rfloor$, are *nested* in $w[C-R\,..\,C+R]$. If a palindrome is even, respectively odd, then its nested palindromes are even, respectively odd.

We write $\langle w\rangle$ for the circular word formed from w. That is, the letters of w are placed in order anticlockwise around the circumference of a circle. The *length* of $\langle w\rangle$ equals the length of w. A *factor* $x[i\,..\,j]$ of a circular word of length n can have $j < i$, in which case $x[i\,..\,j]$ is the concatenation of $x[i\,..\,n]$ and $x[1\,..\,j]$. That is, we are moving anticlockwise around the circle from i to j. We will similarly understand an interval modulo n $[i,j]$ to be $[i,j]$ if $j \ge i$ and $[1,j] \cup [i,n]$ otherwise. Again we can picture the interval as the set of integers in the arc beginning at i moving anticlockwise around the circle to j. We write

$$i \preceq j \preceq k$$

if j is in the arc beginning at i and moving anticlockwise to k. We also define a *set-bounded interval* as follows: If A and B are subsets of $\{1,\dots,n\}$ then

$$[A,B] = \{x : a \preceq x \preceq b \text{ for all } a \in A \text{ and } b \in B\}.$$

A circular word $\langle w\rangle$ of length n is *periodic* with period p if $w[i] = w[i+p]$ (arithmetic modulo n) for all i in $\{1,\dots,n\}$. Clearly n is always a period of $\langle w\rangle$ and if p is the least period of $\langle w\rangle$ then p is a divisor of n. Note that w can be periodic without $\langle w\rangle$ being so, since the periodicity of w only requires $w[i] = w[i+p]$ for $1 \le i \le n-p$. In fact, if $\langle w\rangle$ is the non-periodic circular word $\langle a^k b a^k b a\rangle$, which has length $2k+3$, we have $w[i] = w[i+k+1]$ for $i = 1,\dots,2k+1$. It is not possible to have a non-periodic circular word of length n with $w[i] = w[i+p]$ for all i in $\{1,\dots,n-1\}$.

The following well-known theorem is due to Droubay, Justin, and Pirillo [5]. We give a proof since the ideas here will be used later.

Proposition 5. *The number of distinct non-empty palindromes in a word of length n is at most n.*

Proof. If two palindromes end at the same place then the shorter is a suffix of the longer. It is therefore also a prefix of the longer and so has occurred earlier in the word. Thus at each position there is the end of at most one palindrome making its first appearance in the word. The proposition follows. □

We consider the maximum number of distinct palindromes in a circular word of length n. We only count palindromes of length at most n. Say the maximum number of distinct palindromes in a circular word of length n is $\pi(n)$. We note that in general $\pi(n)$ is larger than n. For example with $\langle aabb\rangle$ we get 6 palindromes: *a*, *b*, *aa*, *bb*, *abba*, *baab*. Table 1 shows values of $\pi(n)$ for low n.

Lemma 6. *For words whose length n is divisible by 3 and at least 9 we have*

$$\pi(n) \ge \frac{5n-6}{3}.$$

Table 1
The first column is the length of the word, the second the maximum number of palindromes and the third an example word attaining the maximum.

n	$\pi(n)$	Example
1	1	$\langle a\rangle$
2	2	$\langle aa\rangle$
3	4	$\langle aab\rangle$
4	6	$\langle aabb\rangle$
5	7	$\langle aaaab\rangle$
6	9	$\langle aaaabb\rangle$
7	10	$\langle aaaaaab\rangle$
8	12	$\langle aaaaaabb\rangle$
9	13	$\langle aaaaaaaab\rangle$
10	15	$\langle aaaaaaaabb\rangle$
11	16	$\langle aaaaaaaaaab\rangle$
12	18	$\langle aaaaaaaaaabb\rangle$
13	19	$\langle aaaaaaaaaaaab\rangle$
14	21	$\langle aaaaaaaaaaaabb\rangle$
15	23	$\langle aaaaabaaaabaaab\rangle$
16	24	$\langle aaaaaaaaaaaaaabb\rangle$
17	26	$\langle aaaaaabaaaaabaaab\rangle$
18	28	$\langle aaaaaabaaaaabaaaab\rangle$
19	29	$\langle aaaaaaabaaaaaabaaab\rangle$
20	31	$\langle aaaaaaabaaaaaabaaaab\rangle$
21	33	$\langle aaaaaaabaaaaaabaaaaab\rangle$

Proof. For $k \geq 1$ the word $\langle a^k ba^{k+1} ba^{k+2} b\rangle$ contains the palindromes

$$\begin{array}{ll} a^i & \text{for } i = 1, \ldots, k+2 \\ a^i ba^i & \text{for } i = 0, \ldots, k+1 \\ a^i ba^k ba^i & \text{for } i = 0, \ldots, k+1 \\ a^i ba^{k+1} ba^i & \text{for } i = 0, \ldots, k \\ a^i ba^{k+2} ba^i & \text{for } i = 0, \ldots, k \end{array}$$

and has length $n = 3k + 6$. The total number of palindromes is therefore

$$5k + 8 = (5n - 6)/3$$

as required. □

We will call the word $\langle a^k ba^{k+1} ba^{k+2} b\rangle$ the kth *Biggles word*.[1]
The main result of this paper is the following:

Theorem 7. *For all n*

$$\pi(n) < 5n/3.$$

Most of the rest of the paper is devoted to proving this theorem. We label the letters of a circular word $\langle w\rangle$ of length n as $w[1], \ldots, w[n]$ as usual and call $w[1]$ the *origin* of the word. This choice is arbitrary since we can relabel the letters of the word to make any of them the origin. We want to count the number of distinct palindromes in $\langle w\rangle$. To do this we put them in two classes: *palindromes of the first kind* have the form $w[i \ldots j]$ with $i \leq j$, so they lie entirely in $w[1 \mathinner{..} n]$ and do not straddle the ends of the word. *Palindromes of the second kind* have the form $w[i \ldots j]$ with $i > j$, so they contain the factor $w[n \mathinner{..} 1]$. Palindromes of the first kind are just those in the linear word w, so by Proposition 5 there are at most $|w|$ of them.

We call a maximal palindrome of length at least $2n/3$, which is not contained in another palindrome of greater length, a *long palindrome*. Below we show that a counterexample to Theorem 7 contains at least one long palindrome. In Corollary 17 we show there are either four or five. In Section 2 we show that no counterexample exists containing exactly four long palindromes and in Section 3 that non exists containing exactly five. This will complete the proof that no counterexample exists.

Lemma 8. *If a counterexample to Theorem 7 exists then it contains at least one long palindrome.*

[1] The word is named for the writer's cat. Biggles is an idle creature and has done nothing to deserve this honour.

Proof. Suppose that $\langle w \rangle$ is a counterexample to Theorem 7. A circular word of length n contains at most n palindromes of the first kind. It must therefore contain at least $2n/3$ palindromes of the second kind which are distinct from all palindromes of the first kind. No two of these palindromes of the second kind can end at the same place, so the longest of these ends at $w[j]$ for some $j \geq 5n/3$ and must straddle the origin and so has length greater than $5n/3$, and so is a long palindrome. $\square$

Lemma 9. *If $P(C_1, R_1)$ and $P(C_2, R_2)$ are palindromes which contain each other's centres and are such that neither is a proper factor of the other then their union has period $2|C_2 - C_1|$.*

Proof. Say the palindromes are in a word w. Without loss of generality suppose that $R_1 \leq R_2$ and $C_1 < C_2$. The proofs of other cases follow from symmetry. Since neither palindrome is a proper factor the other we have

$$C_1 - R_1 \leq C_2 - R_2, \qquad C_1 + R_1 \leq C_2 + R_2 \tag{3}$$

so that $P(C_1, R_1) \cup P(C_2, R_2) = w[C_1 - R_1 .. C_2 + R_2]$. Suppose

$$i \in \big[C_1 - R_1, C_2 + R_2 - 2(C_2 - C_1)\big] = [C_1 - R_1, 2C_1 + R_2 - C_2].$$

From (3)

$$C_1 + R_2 \leq C_2 + R_1$$

so that $w[i]$ is in $P(C_1, R_1)$ and by (1) we have

$$w[i] = w[2C_1 - i].$$

Now

$$2C_1 - i \in \big[2C_1 - (2C_1 + R_2 - C_2), 2C_1 - (C_1 - R_1)\big] = [C_2 - R_2, C_1 + R_1]$$

so, by (3), $w[2C_1 - i]$ is in $P(C_2, R_2)$. By (1) again we have

$$w[2C_1 - i] = w\big[i + 2(C_2 - C_1)\big]$$

as required. $\square$

Lemma 10. *If $\langle w \rangle$ is a counterexample to Theorem 7 then it does not have period less than $|w|$.*

Proof. Say $\langle w \rangle$ has length n and minimum period $p < n$, then p is a proper divisor of n so $p \leq n/2$. If $\langle w \rangle$ is a counterexample then it contains at least $2n/3$ palindromes of the second kind. These begin before $w[1]$ and end at pairwise different positions. The longest must contain $w[\lfloor 2n/3 \rfloor]$. Say this is $P(C, R)$. Note that $P(C - p, R)$ is also a palindrome of the second kind and has appeared before $P(C, R)$. Thus $P(C, R)$ equals a palindrome that has already been counted and $\langle w \rangle$ is not a counterexample. $\square$

We say that two long palindromes with centres C_1 and C_2 are *adjacent* if there is no long palindrome with its centre lying between C_1 and C_2.

Lemma 11. *If $P(C_i, R_i)$ and $P(C_{i+1}, R_{i+1})$ are adjacent long palindromes in a circular word $\langle w \rangle$ with $C_i < C_{i+1}$ then*

$$R_i + R_{i+1} \geq 2n/3 + C_{i+1} - C_i - 1, \tag{4}$$

and if $P(C_1, R_1)$ and $P(C_t, R_t)$ are adjacent long palindromes with C_1 and C_t on opposite sides of the origin then

$$R_1 + R_t \geq 2n/3 + C_1 + n - C_t - 1. \tag{5}$$

Proof. Take the conjugate of w for which

$$C_{i+1} - R_{i+1} = 1$$

so that $P(C_{i+1}, R_{i+1})$ is of the first kind. So $w[2n/3]$ is covered by $P(C_i, R_i)$, that is

$$C_i + R_i \geq 2n/3$$

and so

$$R_i + R_{i+1} \geq 2n/3 + C_{i+1} - 1 - C_i.$$

This establishes (4). Inequality (5) can be proved in the same way. $\square$

Proposition 12. *If $P(C_1, R_1), \dots, P(C_t, R_t)$ are the long palindromes in a counterexample to Theorem 7 then*

$$2\sum_{i=1}^{t} R_i > 2tn/3 + n - t. \tag{6}$$

Proof. We assume, without loss of generality, that $0 < C_1 < \dots < C_t$. Then summing (4) over i and adding (5) gives

$$2\sum_{i=1}^{t} R_i > 2tn/3 - t + \sum_{i=1}^{t}(C_{i+1} - C_i) + n.$$

We note that

$$\sum_{i=1}^{t}(C_{i+1} - C_i) = 0,$$

giving (6). □

Proposition 13. *If w is a palindrome of length n then the number of distinct palindromes with length at most n in the circular word $\langle w\rangle$ is at most $\lceil \frac{3n-1}{2} \rceil$, and this bound is best possible.*

Proof. Let $v = ww$. Any palindrome in $\langle w\rangle$ will occur in v, and any palindrome with length at most n that occurs in v will occur in $\langle w\rangle$. So the number of distinct palindromes in $\langle w\rangle$ equals the number of distinct palindromes in v minus the number of distinct palindromes in v with length greater than n.

By Proposition 5 the number of distinct palindromes in v is at most $2n$.

The set of palindromes in v with length greater than n contains the set $\{v[i \mathinner{..} 2n + 1 - i], i = 1, \dots, \lfloor (n+1)/2 \rfloor\}$ which has cardinality $\lfloor (n+1)/2 \rfloor$.

Therefore the number of distinct palindromes in $\langle w\rangle$ is at most

$$2n - \left\lfloor \frac{n+1}{2} \right\rfloor = \left\lceil \frac{3n-1}{2} \right\rceil.$$

This bound is attained by the circular words $\langle a^k b a^k\rangle$ for $k \geq 2$. □

Proposition 14. *If $\langle w\rangle$ is a counterexample to Theorem 7 then no conjugate of w is a palindrome, hence any palindrome in $\langle w\rangle$ has length at most $n-1$.*

Proof. Suppose w has a conjugate v which is a long palindrome. Then $\langle v\rangle$, which equals $\langle w\rangle$, contains less than $5n/3$ palindromes by Proposition 13. Therefore $\langle w\rangle$ is not a counterexample. □

Corollary 15. *If w is a counterexample to Theorem 7 and $P(C, R)$ is a long palindrome in $\langle w\rangle$ then*

$$n/3 \leq R \leq n/2 - 1.$$

Proof. The left hand side of the inequality comes from the definition of a long palindrome. For the right hand side note that the length of $P(C, R)$ is $2R + 1$. By Proposition 14 this is strictly less than n which implies the result. □

Proposition 16. *If w is a counterexample to Theorem 7 with long palindromes $\{P(C_i, R_i) : i = 1, \dots, t\}$ where $1 \leq C_1 < \dots < C_t \leq n$ then for $i = 1, \dots, t-1$ we have*

$$n/6 < C_{i+1} - C_i \leq n/3 - 1 \tag{7}$$

and

$$n/6 < C_1 + n - C_t \leq n/3 - 1. \tag{8}$$

Proof. Consider i such that $1 \leq i < n$. By Lemma 11 we have

$$R_i + R_{i+1} \geq 2n/3 + C_{i+1} - C_i - 1$$

and by Corollary 15 we have

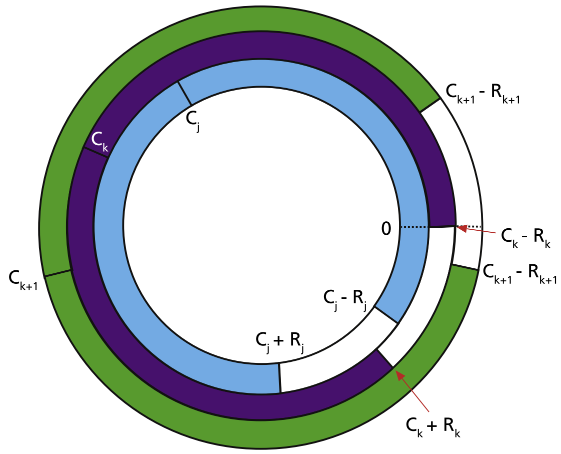

Fig. 1. The situation considered in the proof of Proposition 16.

$$n-2 \geq R_i + R_{i+1}.$$

Together these give

$$n/3-1 \geq C_{i+1} - C_i$$

which is the right hand side of (7). For (8) we just consider a suitable conjugate of w (see Fig. 1).

For the other side suppose that the minimum value of $C_{i+1} - C_i$ is attained with $i = k$. The gaps between consecutive centres of long palindromes cannot all be equal as it is easily shown with two applications of (1) that the whole of $\langle w \rangle$ would then be periodic. This is impossible by Lemma 10. We can therefore assume, without loss of generality, that $C_k - (C_{k+1} - C_k)$ is not the centre of a long palindrome. We consider that conjugate of w for which

$$C_k - R_k = 1 \tag{9}$$

so that $P(C_k, R_k)$ is not of the second kind. Since $\langle w \rangle$ is a counterexample to Theorem 7 it contains a long palindrome of the second kind which contains $w[\lceil 2n/3 \rceil]$. Suppose this is $P(C_{k+1}, R_{k+1})$. Then

$$C_{k+1} - R_{k+1} \leq 0 < C_k - R_k$$

which would imply that $P(C_k, R_k)$ is entirely contained in $P(C_{k+1}, R_{k+1})$ which contradicts the definition of a long palindrome. We conclude that neither $P(C_k, R_k)$ nor $P(C_{k+1}, R_{k+1})$ is of the second kind. But $w[\lceil 2n/3 \rceil]$ is covered by some palindrome of the second kind. Say this is $P(C_j, R_j)$. Then we have $C_j - R_j \leq 0$ and $C_j + R_j \geq \lceil 2n/3 \rceil$. With Corollary 15 we have

$$C_j \geq \lfloor 2n/3 \rfloor - n/2 + 1 > n/6. \tag{10}$$

Suppose, for the sake of contradiction, that

$$C_{k+1} - C_k \leq C_j. \tag{11}$$

By Proposition 9 $P(C_k, R_k) \cup P(C_{k+1}, R_{k+1})$ has period $2(C_{k+1} - C_k)$ and $P(C_j, R_j) \cup P(C_k, R_k)$ has period $2(C_k - C_j)$. Their intersection, which is $P(C_k, R_k)$, has both periods. $P(C_k, R_k)$ has length $2R_k + 1$ which, by (9), equals $2C_k - 1$. By (11) this is at least

$$2(C_{k+1} - C_j) - 1 = 2(C_{k+1} - C_k) + 2(C_k - C_j) - 1.$$

By Lemma 1 $P(C_k, R_k)$ therefore has period

$$\Delta = \gcd\bigl(2(C_{k+1} - C_k), 2(C_k - C_j)\bigr)$$

and by Lemma 3 this period extends to the whole of

$$P(C_j, R_j) \cup P(C_k, R_k) \cup P(C_{k+1}, R_{k+1}) = w[C_j - R_j, C_{k+1} + R_{k+1}]$$

which we will call X. We note that $2(C_{k+1} - C_k) \geq \Delta$. We consider two cases depending on the sign of $R_{k+1} - R_j$.

Case 1: $R_{k+1} - R_j \geq 0$. Let $q = \lceil (C_{k+1} - C_j)/\Delta \rceil - 1$ so that

$$C_{k+1} - \Delta \leq C_j + q\Delta < C_{k+1}.$$

It follows that $C_j + q\Delta + R_j \leq C_{k+1} + R_{k+1}$ so that

$$w[C_j + q\Delta - R_j \mathinner{..} C_j + q\Delta + R_j]$$

is inside X and, since X has period Δ, is a copy of $P(C_j, R_j)$. Thus $P(C_j + q\Delta, R_j)$ is a long palindrome with its centre, $C_j + q\Delta$, in the interval $[C_{k+1} - \Delta, C_{k+1})$. Since $\Delta \leq 2(C_{k+1} - C_k)$ this is a subinterval of

$$\big[C_{k+1} - 2(C_{k+1} - C_k), C_{k+1}\big] = \big[C_k - (C_{k+1} - C_k), C_{k+1}\big].$$

By our earlier assumption that $C_k - (C_{k+1} - C_k)$ is not the centre of a long palindrome $C_j + q\Delta$ is in the interior of the interval, so we have a long palindrome whose centre is at a distance from C_k less than $C_{k+1} - C_k$, contradicting the minimality of this distance. We conclude that Case 1 is impossible.

Case 2: $R_{k+1} - R_j < 0$. We note that, from the definition of Δ,

$$\Delta \leq (C_{k+1} - C_k) + (C_k - C_j)$$

so that

$$C_{k+1} - R_{k+1} - \Delta = C_j + (C_{k+1} - C_k) + (C_k - C_j) - R_{k+1} - \Delta \geq C_j - R_j.$$

Thus $w[C_{k+1} - R_{k+1} - \Delta \mathinner{..} C_{k+1} + R_{k+1} - \Delta]$ is in X, and so is a long palindrome by the Δ-periodicity of X. Since $\Delta \leq 2(C_{k+1} - C_k)$ its centre, which is $C_{k+1} - \Delta$, is in the interval $[C_k - (C_{k+1} - C_k), C_{k+1})$ which is impossible as in Case 1.

We conclude that our original assumption (11) is false, so that, by (10), $C_{k+1} - C_k > n/6$. Since $C_{k+1} - C_k$ is the minimum gap between consecutive centres we are done. □

Corollary 17. *If a counterexample to Theorem 7 exists then it contains either exactly 4 or exactly 5 long palindromes.*

Proof. If a counterexample contained 3 or fewer long palindromes then there would be adjacent centres at least $n/3$ apart and if it contained 6 or more there would be centres at most $n/6$ apart. Either case would contradict Proposition 16. □

2. Four long palindromes

In this section we show that $\langle w \rangle$ cannot be a counterexample to Theorem 7 if it contains exactly four long palindromes. In the following propositions we use set-bounded intervals. These were defined in the introduction.

Proposition 18. *Suppose that $\langle w \rangle$ is a counterexample of length n with exactly four long palindromes $P(C_1, R_1)$, $P(C_2, R_2)$, $P(C_3, R_3)$ and $P(C_4, R_4)$ where $1 \leq C_1 < C_2 < C_3 < C_4 \leq n$. Then $\langle w \rangle$ has period $2(C_2 - C_1 + C_4 - C_3)$.*

Proof. We write p for $2(C_2 - C_1 + C_4 - C_3)$ and I_1 for the set-bounded interval

$$\big[\{C_1 - R_1, C_3 - 2(C_2 - C_1) - R_3\}, \{2C_1 - C_2 + R_2, 2(C_3 - C_2 + C_1) - C_4 + R_4\}\big]$$

We will show that for $x \in I_1$ we have

$$w[x] = w\big[2(C_2 - C_1 + C_4 - C_3) + x\big].$$

We fix $x \in I_1$. Then $x \geq C_1 - R_1$ and using Corollary 15 and Proposition 16

$$\begin{aligned} x &\leq C_1 - (C_2 - C_1) + R_2 \\ &< C_1 - n/6 + n/2 \\ &< C_1 + R_1 \end{aligned} \tag{12}$$

so $x \in P(C_1, R_1)$. Therefore $w[x] = w[2C_1 - x]$. We now have $2C_1 - x$ in the interval

$$\big[\{C_2 - R_2, C_4 - 2(C_3 - C_2) - R_4\}, \{C_1 + R_1, 2C_2 - C_3 + R_3\}\big]$$

which, by reasoning similar to (12), is inside $P(C_2, R_2)$. We therefore have

$$w[2C_1 - x] = w\big[2(C_2 - C_1) + x\big].$$

We now have $2(C_2 - C_1) + x$ in the interval

$$\big[\{2C_2 - C_1 - R_1, C_3 - R_3\}, \{C_2 + R_2, R_4 - C_4 + 2C_3\}\big]$$

which, by reasoning similar to (12), is inside $P(C_3, R_3)$. We therefore have $w[2(C_2 - C_1) + x] = w[2(C_3 - C_2 + C_1) - x]$. We now have $2(C_3 - C_2 + C_1) - x$ in the interval

$$\big[\{2C_3 - C_2 - R_2, C_4 - R_4\}, \{C_1 + 2C_3 - 2C_2 + R_1, C_3 + R_3\}\big].$$

By Proposition 16 $C_4 > C_3 + n/6$ and by Corollary 15 $R_3 - R_4 < n/6 - 1$. Hence

$$C_3 + R_3 < C_4 - n/6 + R_4 + n/6 - 1 < C_4 + R_4,$$

and the interval above is inside $P(C_4, R_4)$. We therefore have $w[2(C_3 - C_2 + C_1) - x] = w[2(C_4 - C_3 + C_2 - C_1) + x]$, so that for any x in I_1 we have

$$w[x] = w\big[x + 2(C_4 - C_3 + C_2 - C_1)\big] = w[x + p]$$

as required.

Similarly we can show that if x is in the interval

$$\big[\big\{C_2 - R_2, C_4 - 2(C_3 - C_2) - R_4\big\}, \big\{2C_2 - C_3 + R_3, 2(C_4 - C_3 + C_2) - C_1 + R_1\big\}\big]$$

then

$$w[x] = w\big[2(C_3 - C_2 + C_1 - C_4 + n) + x\big] = w[x - p].$$

Then $x - p$ lies in the interval

$$\big[\big\{2C_1 - C_4 - R_4, 2(C_1 - C_4 + C_3) - C_2 - R_2\big\}, \big\{C_1 + R_1, C_3 - 2(C_4 - C_1) + R_3\big\}\big]$$

which we call I_4. It follows that if x is in I_4 then

$$w[x] = w\big[x - 2(C_1 - C_4 + C_3 - C_2)\big] = w\big[x + 2(C_2 - C_1 + C_4 - C_3)\big] = w[x + p].$$

Similarly we have $w[x] = w[x + p]$ for x in the intervals

$$\big[\big\{2C_3 - C_2 - R_2, 2(C_1 - C_2 + C_3) - C_4 - R_4\big\}, \big\{C_3 + R_3, C_1 + 2(C_3 - C_2) + R_1\big\}\big]$$

and

$$\big[\big\{C_3 - R_3, C_1 - 2(C_4 - C_3) - R_1\big\}, \big\{2C_3 - C_4 + R_4, 2(C_1 - C_4 + C_3) - C_2 + R_2\big\}\big]$$

which we call, respectively, I_2 and I_3. The I_3 case is similar to I_1 and the I_2 case is similar to I_4. We must now show that

$$I_1 \cup I_2 \cup I_3 \cup I_4 = \langle w \rangle.$$

We do this by first finding points P_1, P_2, P_3 and P_4, with

$$P_1 > P_2 > P_3 > P_4 > P_1 - n, \tag{13}$$

which lie in the intervals I_1, I_2, I_3 and I_4 respectively; then show that, for $1 \le i \le 4$ the lower bound of I_{i+1} is less than the upper bound of I_i with arithmetic on the indices modulo 4.

We set $P_1 = 2C_1 - C_2$ and show that it lies in I_1 which is the set-bounded interval

$$\big[\big\{C_1 - R_1, C_3 - 2(C_2 - C_1) - R_3\big\}, \big\{2C_1 - C_2 + R_2, 2(C_3 - C_2 + C_1) - C_4 + R_4\big\}\big].$$

We have

$$P_1 - (C_1 - R_1) = R_1 - (C_2 - C_1)$$
$$P_1 - \big(C_3 - 2(C_2 - C_1) - R_3\big) = R_3 - (C_3 - C_2).$$

Using Proposition 16 and Proposition 15 we see that each of these lies in the interval $(0, n/3)$. For the upper bound

$$2C_1 - C_2 + R_2 - P_1 = R_2$$
$$2(C_3 - C_2 + C_1) - C_4 + R_4 - P_1 = R_4 + (C_3 - C_2) - (C_4 - C_3).$$

Using Proposition 16 and Corollary 15 we see that each of these lies in the interval $(n/6, 2n/3)$. Hence P_1 is inside I_1. Similarly we can show that $2C_2 - C_3$ is inside the interval I_2':

$$\left[\left\{C_2 - R_2, C_4 - 2(C_3 - C_2) - R_4\right\}, \left\{2C_2 - C_3 + R_3, 2(C_4 - C_3 + C_2) - C_1 + R_1\right\}\right].$$

Recalling the construction of I_4 we see that

$$(2C_2 - C_3) - p = 2C_1 - 2C_4 + C_3$$

is inside I_4. We call this point P_4. P_3 and P_2 are found in the same ways as P_1 and P_4 respectively, giving:

$$P_1 = 2C_1 - C_2$$
$$P_2 = C_1 - 2C_2 + 2C_3$$
$$P_3 = 2C_3 - C_4$$
$$P_4 = C_3 - 2C_4 + 2C_1.$$

We now establish (13). We have

$$P_1 - P_2 \equiv (C_1 + n - C_4) + (C_4 - C_3) - (C_3 - C_2) \pmod n$$
$$P_2 - P_3 \equiv (C_4 - C_3) + (C_3 - C_2) - (C_2 - C_1) \pmod n$$
$$P_3 - P_4 \equiv (C_3 - C_2) + (C_2 - C_1) - (C_1 + n - C_4) \pmod n$$
$$P_4 - P_1 \equiv (C_2 - C_1) + (C_1 + n - C_4) - (C_4 - C_3) \pmod n.$$

Using Proposition 16 we find that the right hand side of each congruence is in the interval $(0, n/2)$ which implies (13).

We now show that interval I_1 has non-empty intersection with interval I_2. We will assume that

$$R_i = \lceil n/3 \rceil \quad \text{for } i = 1, \ldots, 4 \tag{14}$$

This involves no loss of generality since increasing R_i can only lengthen the intervals. The assumption will avoid some complications. We write L_1 for the lower bound of I_1 and U_2 for the upper bound of I_2. Thus

$$L_1 = C_1 - R_1 \text{ or } C_3 - 2(C_2 - C_1) - R_3$$
$$U_2 = C_3 + R_3 \text{ or } C_1 + 2(C_3 - C_2) + R_1. \tag{15}$$

If I_1 and I_2 do not intersect then

$$P_1 > L_1 > U_2 > P_2. \tag{16}$$

As noted above $P_1 - P_2 < n/2$, so a necessary condition for (16) is that $L_1 - U_2 < n/2$, which is equivalent to

$$U_2 - L_1 \geq n/2. \tag{17}$$

There are three possible values for $U_2 - L_1$ depending on which alternatives occur in (15) (two of the four combinations result in the same value). These are

$$C_3 - C_1 + R_1 + R_3 = (C_3 - C_2) + (C_2 - C_1) + 2\lceil n/3 \rceil$$
$$2(C_3 - C_2) + 2R_1 = 2(C_3 - C_2) + 2\lceil n/3 \rceil$$
$$2(C_2 - C_1) + 2R_3 = 2(C_2 - C_1) + 2\lceil n/3 \rceil.$$

By Proposition 16 and (14) each of these is in the interval $(n, 4n/3) \equiv (0, n/3) \pmod n$, contradicting (17). This shows that I_1 and I_2 do indeed intersect. Similar arguments show that each of $I_2 \cap I_3$, $I_3 \cap I_4$ and $I_4 \cap I_1$ is non-empty. This shows that

$$I_1 \cup I_2 \cup I_3 \cup I_4 = \langle w \rangle,$$

as required. The Proposition follows. □

Lemma 19. *Let $P(C_1, R_1)$, $P(C_2, R_2)$ and $P(C_3, R_3)$ be palindromes in a word w with $C_2 - C_1 = x + y$, $C_3 - C_2 = x$, where x and y are positive, $R_i > x \geq y$ for $i = 1, \ldots, 3$,*

$$R_3 \geq R_2 - x \tag{18}$$

and

$$R_1 \geq R_2 - x + y. \tag{19}$$

Then each of the palindromes $P(C_3, R_2 - x)$ and $P(C_1, R_2 - x + y)$ has period $2y$.

Proof. Note that $C_2 = C_1 + x + y$ and $C_3 = C_1 + 2x + y$. We write p for $w[2C_3 - C_2 - R_2 .. C_2 + R_2]$ which is the palindrome $P(C_3, R_2 - x) = w[C_1 + 3x + y - R_2 .. C_1 + x + y + R_2]$. Note that $|p| = 2R_2 - 2x + 1$. The reversal of p in $P(C_2, R_2)$ is

$$\begin{aligned} &w\big[2C_2 - (C_3 + R_2 - x) .. 2C_2 - (C_3 - R_2 + x)\big] \\ &\quad = w\big[2(C_1 + x + y) - (C_1 + x + y + R_2) .. 2(C_1 + x + y) - (C_1 + 3x + y - R_2)\big] \\ &\quad = w[C_1 + x + y - R_2 .. C_1 - x + y + R_2]. \end{aligned}$$

This equals p since p is a palindrome and so is $P(C_1 + y, R_2 - x)$. By (19) it lies inside $P(C_1, R_1)$. Its reversal in $P(C_1, R_1)$ is

$$w\big[2C_1 - (C_1 - x + y + R_2) .. 2C_1 - (C_1 + x + y - R_2)\big] = w[C_1 + x - y - R_2 .. C_1 - x - y + R_2]$$

which again equals p. Thus p is a border of the word $w[C_1 + x - y - R_2 .. C_1 - x + y + R_2]$, which is the palindrome $P(C_1, R_2 - x + y)$. By Lemma 2 this palindrome therefore has period

$$\big|w[C_1 + x - y - R_2 .. C_1 - x + y + R_2]\big| - |p| = (2R_2 - 2x + 2y + 1) - (2R_2 - 2x + 1) = 2y,$$

as required. Since p is a factor of this word it also has period $2y$; that is, $P(C_3, R_2 - x)$ has period $2y$. □

Corollary 20. *There is no counter-example to Theorem 7 with exactly four long palindromes.*

Proof. Suppose that $\langle w\rangle$ is a counterexample to Theorem 7 and that it contains exactly four long palindromes. Then by Proposition 18, and using the notation of that proposition, $\langle w\rangle$ has period $2(C_2 - C_1 + C_4 - C_3)$. It also has period n. If $\gcd(n, 2(C_2 - C_1 + C_4 - C_3)) < n$ then $\langle w\rangle$ is periodic, and so, by Lemma 10, cannot be a counter-example. We conclude that the greatest common divisor is n. Clearly $C_2 - C_1 + C_4 - C_3 < n$ so we must have

$$C_2 - C_1 + C_4 - C_3 = n/2. \qquad (20)$$

Since

$$(C_4 - C_3) + (C_3 - C_2) + (C_2 - C_1) + (C_1 + n - C_4) = n$$

we have

$$(C_3 - C_2) + (C_1 + n - C_4) = n/2.$$

Suppose, without loss of generality, that $C_3 - C_2$ is the smallest inter-centre gap and let $C_3 - C_2 = x$. Then we have

$$\begin{aligned} C_2 - C_1 &= x + y \\ C_3 - C_2 &= x \\ C_4 - C_3 &= x + z \\ C_1 + n - C_4 &= x + y + z, \end{aligned} \qquad (21)$$

for some non-negative numbers y and z so that

$$n = 4x + 2y + 2z. \qquad (22)$$

The structure of the word is shown on the left of Fig. 2. Note that $R_2 - x < n/2 - n/6 \le R_3$ by Proposition 16 and Corollary 15 so (18) holds. Let $R_1' = \min\{R_1, R_2 - x + y\}$. Then $P(C_1, R_1')$ is a palindrome which can play the role of $P(C_1, R_1)$ in Lemma 19. We see that (19) is satisfied. Thus $P(C_3, R_2 - x)$ and $P(C_1, R_2 - x + y)$ each have period $2y$. Call these palindromes Y and Y' respectively. We also apply this lemma to the reverse of $\langle w\rangle$ with C_2 playing the role of C_1 and C_4 the role of C_3. Then $P(C_2, R_3 - x)$ and $P(C_4, R_3 - x + z)$ each have period $2z$. Call these palindromes Z and Z' respectively. The elaborated structure of the word is shown on the right of Fig. 2. The intersection of Y and Z is named θ. This has length

$$C_3 + R_2 - x - (C_4 - R_3 + x - z) + 1 = R_2 + R_3 - 2x + z - (C_4 - C_3) + 1 = R_2 + R_3 - 3x + 1.$$

Similar calculations show that each of $Z \cap Y'$, $Y' \cap Z'$ and $Z' \cap Y$ has the same length. By palindromicity each intersection is the reverse of the two adjacent intersections so they equal $\bar{\theta}$, θ and $\bar{\theta}$ respectively. By Lemma 11 $|\theta|$ is at least $2n/3 - 2x$ and by Proposition 16 and (21)

$$n/3 - 1 \ge C_1 + n - C4 = x + y + z.$$

Therefore $|\theta| \ge 2y + 2z + 2$. Since θ is the intersection of Y and Z it has periods $2y$ and $2z$. By Lemma 1 it therefore has period $\gcd(2y, 2z)$ and by Lemma 3 this periodicity extends to the whole of $\langle w\rangle$, contradicting Lemma 10. We conclude that $\langle w\rangle$ is not a counterexample to Theorem 7 and so no counterexample with exactly four long palindromes exists. □

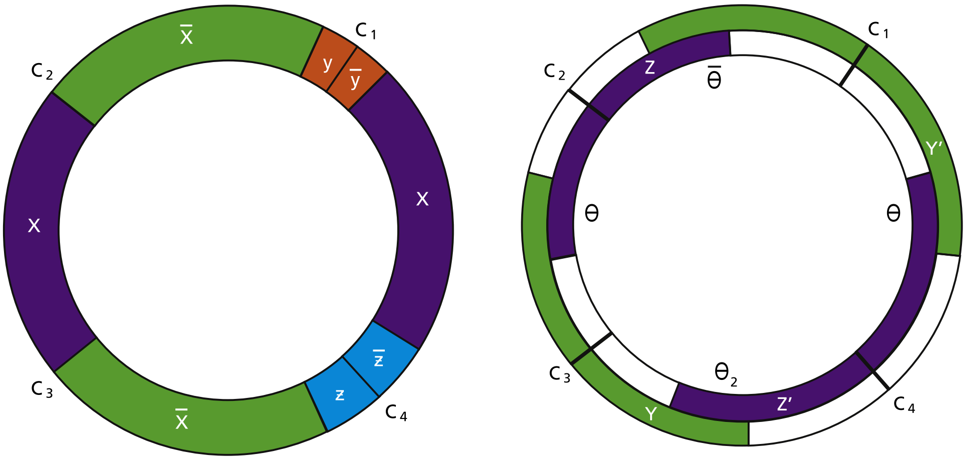

Fig. 2. Four long palindromes. The diagram on the left shows the arrangement of the centres of the four long palindromes, with the location of the centres satisfying (20). The diagram on the right shows the palindromes Y, Y', Z and Z' used in the second part of the proof. Their pair-wise intersections are labelled θ or $\bar{\theta}$, $\bar{\theta}$ being the reverse of θ.

3. Five long palindromes

The remaining case to consider is $\langle w \rangle$ containing exactly five long palindromes.

Lemma 21. *There is no counter-example to Theorem 7 with exactly five long palindromes.*

Proof. We show that if $\langle w \rangle$ contains exactly five long palindromes then it is periodic so, by Lemma 10, not a counterexample. Suppose $\langle w \rangle$ contains $P(C_1, R_1), \ldots, P(C_5, R_5)$. Using Proposition 12 we have

$$\sum_{i=2}^{4}\{2R_i + 1 - 2(C_{i+1} - C_{i-1})\} + \{2R_5 + 1 - 2(C_1 - C_4 + n)\} + \{2R_1 + 1 - 2(C_2 - C_5 + n)\}$$
$$= 2\sum_{i=1}^{5} R_i - 4n + 5$$
$$> n/3.$$

Hence at least one of the terms in parentheses is positive. Without loss of generality suppose $2R_2 + 1 - 2(C_3 - C_1) > 0$, that is:

$$2R_2 + 1 - 2(C_3 - C_2) - 2(C_2 - C_1) > 0. \tag{23}$$

By Proposition 9, $P(C_1, R_1) \cup P(C_2, R_2)$ has period $2(C_2 - C_1)$ and $P(C_2, R_2) \cup P(C_3, R_3)$ has period $2(C_3 - C_2)$. Their intersection, which is $P(C_2, R_2)$, has both periods. $P(C_2, R_2)$ has length $2R_2 + 1$ which, by (23), is at least the sum of the periods less one. By Lemma 1 $P(C_2, R_2)$ therefore has period $p = \gcd(2(C_2 - C_1), 2(C_3 - C_2))$. By Lemma 3 this periodicity extends to the whole of $P(C_1, R_1) \cup P(C_2, R_2) \cup P(C_3, R_3) = w[C_1 - R_1 \mathinner{..} C_3 + R_3]$ which has length $R_1 + C_3 - C_1 + R_3$. This is greater than n but, as noted in the introduction, this is not sufficient to imply that $\langle w \rangle$ is periodic.

First suppose that $C_2 - C_1 = C_3 - C_2$ so that, by Proposition 16,

$$n/3 < p = 2(C_2 - C_1) < 2n/3. \tag{24}$$

We show that the whole of $\langle w \rangle$ is the palindrome $P(C_2, p + \lceil n/3 \rceil)$. To do this we must show that $w[i] = w[2C_2 - i]$ for all i in $[C_2 - (p + \lceil n/3 \rceil), C_2 + (p + \lceil n/3 \rceil)]$. This is immediate if i is in $[C_2 - R_2, C_2 + R_2]$ since this is $P(C_2, R_2)$. Suppose then that i is in $[C_2 + R_2, C_2 + p + \lceil n/3 \rceil]$. By the p-periodicity $w[i] = w[i - p]$ which is in $[C_2 + R_2 - p, C_2 + \lceil n/3 \rceil]$. By (24) this is inside $P(C_2, R_2)$ so we have $w[i] = w[2C_2 - i + p]$. By the p-periodicity again we have $w[i] = w[2C_2 - i]$ as required. Thus $P(C_2, p + \lceil n/3 \rceil)$ is a palindrome and its length is $2p + 2(\lceil n/3 \rceil)$ which is greater than n, so $\langle w \rangle$ is a palindrome, and by Proposition 14 cannot be a counterexample. We conclude that $C_2 - C_1 \neq C_3 - C_2$.

Now suppose, without loss of generality, that $C_3 - C_2 > C_2 - C_1$. Then

$$p = \gcd\bigl(2(C_2 - C_1), 2(C_3 - C_2)\bigr) < (C_3 - C_2) + (C_2 - C_1) = C_3 - C_1.$$

It may be that $p = C_2 - C_1$ or $p = C_3 - C_2$ but these cannot both be true.

Suppose that $C_3 - C_2 \neq p$. Then by the p-periodicity $\langle w\rangle$ contains the palindrome $P(C_3 - p, R')$ where

$$R' = \min\{R_3, C_3 - p - (C_1 - R_1)\}.$$

Clearly its centre lies between C_1 and C_3. If it were C_2 we'd have $C_3 - C_2 = p$, contrary to our assumption. So it is not nested in any of our five long palindromes. However $R' \geq \min\{R_1, R_3\}$ so the palindrome is long, contradicting the hypothesis that $\langle w\rangle$ contains exactly 5 long palindromes.

On the other hand, if $C_2 - C_1 \neq p$ we can consider the palindrome $P(C_1 + p, R'')$ where

$$R'' = \min\{R_1, C_3 + R_3 - (C_1 + p)\}$$

and obtain a contradiction as in the previous case.

Since neither case is possible, the proof is complete. □

Proof of Theorem 7. This is immediate from Corollary 17, Corollary 20 and Proposition 21. □

Open Questions. (1) A non-circular word w of length n is called *rich* if it contains n non-empty palindromes. The Biggles words are rich. What is the maximum number of palindromes in a circular word $\langle w\rangle$ if w is not rich? Much of the paper was concerned with long palindromes. We could also ask for the maximum number of long palindromes in a word of length n, or just for the maximum number of palindromes with length at least αn.

(2) The proof in this paper is extraordinarily long. Surely there is a simpler way to prove Theorem 7. Also, it is likely that the Biggles word is optimal for n congruent to 0 modulo 3, and that if n is congruent to 1 or 2 modulo 3 then the words $a^k b^{k+1} b a^{k+3}$ and $a^k b^{k+2} b a^{k+3}$ respectively are optimal. Can this be proved? The class of such optimal words could be called *circularly rich words*.

(3) In this paper we have relied on the connection between palindromes and periodicity, initially in Proposition 9. Other results concerning the connections between these concepts appear in [2,4,5,8,1]. A *run* in a word is a periodic factor whose length is at least twice the period and which cannot be extended to the left or right without altering the period. It's conjectured that the number of runs in a word of length n is at most n, see [3]. Proposition 9 can be sharpened to say that if the palindromes are maximal and their union has length at least $4|C_2 - C_1|$ then it is a run. The proposition also has the following "semi-converse".

Theorem 22. *If a word w contains a factor $f = w[a..b]$ which has period p, and this factor contains the palindrome $P(C, R_1)$, where*

$$R_1 \geq p/2, \tag{25}$$

then f also contains the palindrome $P(C + p/2, R_2)$ where

$$R_2 = \min\{C + p/2 - a, b - (C + p/2)\}.$$

Proof. Choose i from the interval

$$[\lfloor C + p/2\rfloor, C + p/2 + R_2],$$

and set

$$k = \lfloor (i - C)/p + 1/2\rfloor$$

so that

$$(i - C)/p - 1/2 < k \leq (i - C)/p + 1/2.$$

Therefore, using (25),

$$C + R_1 > i - kp \geq C - R_1$$

so that $w[i - kp]$ is in $P(C, R_1)$ and therefore in f. By periodicity $w[i] = w[i - kp]$ and by palindromicity $w[i - kp] = w[2C + kp - i]$. Now

$$2C + p - i \geq 2C + p - (C + p/2 + R_2) = C + p/2 - R_2 \geq C + p/2 - (C + p/2 - a) = a$$

so both $2C + kp - i$ and $2C + p - i$ are in f. Therefore $w[2C + p - i] = w[2C + kp - i]$. Collecting the equations we have $w[i] = w[2C + p - i]$ for all i in $[\lfloor C + p/2\rfloor, C + p/2 + R_2]$, from which it follows that $P(C + p/2, R_2)$ is a palindrome. □

Periodicity is fundamental to the combinatorics of words and the connections between periodicity and palindromicity merit further investigation.

Acknowledgements

I thank Julie Caddy and Greg Gamble for their assistance. I also sincerely thank the referees for their very thorough refereeing of this paper. It must have been a tedious task.

References

[1] A. Blondin-Massé, S. Brlek, A. Garon, S. Labbé, Equations on palindromes and circular words, Theoret. Comput. Sci. 412 (2011) 2922–2930.
[2] S. Brlek, S. Hamel, M. Nivat, C. Reutenauer, On the palindromic complexity of infinite words, Internat. J. Found. Comput. Sci. 15 (2004) 293–306.
[3] M. Crochemore, L. Ilie, L. Tinta, The "runs" conjecture, Theoret. Comput. Sci. 412 (2011) 2931–2941.
[4] A. De Luca, Sturmian words: structure, combinatorics, and their arithmetics, Theoret. Comput. Sci. 183 (1) (1997) 45–82.
[5] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, Theoret. Comput. Sci. 255 (2001) 539–553.
[6] N.J. Fine, H.S. Wilf, Uniqueness theorem for periodic functions, Proc. Amer. Math. Soc. 16 (1965) 109–114.
[7] A.S. Fraenkel, J. Simpson, An extension of the periodicity lemma to longer periods (invited lecture), in: Combinatorial Pattern Matching, Springer, Berlin, Heidelberg, 2001.
[8] M.G. Castelli, F. Mignosi, A. Restivo, Fine and Wilf's theorem for three periods and a generalization of Sturmian words, Theoret. Comput. Sci. 218 (1999) 83–94.
[9] M. Lothaire, Algebraic Combinatorics on Words, Cambridge University Press, 2002.