A characterization of balanced episturmian sequences

Geneviève Paquin*† Laurent Vuillon‡

Submitted: Nov 21, 2006; Accepted: Apr 24, 2007; Published: May 9, 2007
Mathematics Subject Classification: 68R15

Abstract

It is well-known that Sturmian sequences are the non ultimately periodic sequences that are balanced over a 2-letter alphabet. They are also characterized by their complexity: they have exactly $(n+1)$ distinct factors of length n. A natural generalization of Sturmian sequences is the set of infinite episturmian sequences. These sequences are not necessarily balanced over a k-letter alphabet, nor are they necessarily aperiodic. In this paper, we characterize balanced episturmian sequences, periodic or not, and prove Fraenkel's conjecture for the special case of episturmian sequences. It appears that balanced episturmian sequences are all ultimately periodic and they can be classified in 3 families.

1 Introduction

Sturmian sequences are exactly the non ultimately periodic balanced sequences over a 2-letter alphabet [6, 18]. A sequence s is balanced if for every letter a, the number of a's in any two n-length factors differs by at most 1, for any n. Sturmian sequences are also characterized by their number of n-length factors: they always have $(n+1)$ factors of length n, for every n. For Sturmian sequences, the two conditions are equivalent. There are two different generalizations of Sturmian sequences for alphabets of cardinality $k \geq 3$. A natural generalization of Sturmian sequences is the set of infinite episturmian sequences. It was first obtained by a construction due to de Luca [17] which uses the palindromic closure. The class of strict episturmian sequences over a 3-letter alphabet also appears in [19] and is studied in [3]. The set of episturmian sequences have been extensively studied by Droubay, Justin and Pirillo [9, 15, 16] more recently. The second generalization of Sturmian sequences is the set of balanced sequences studied in [5, 22, 23].

*with the support of NSERC (Canada)

†Laboratoire de combinatoire et d'informatique mathématique, Université du Québec à Montréal, CP. 8888 Succ. Centre-Ville, Montréal, (QC) CANADA, H3C 3P8, paquin@lacim.uqam.ca

‡Laboratoire de mathématiques, CNRS UMR 5127, Université de Savoie, 73376 Le Bourget-du-lac cedex, France, Laurent.Vuillon@univ-savoie.fr

It is interesting to note that the two notions coincide for Sturmian sequences, which are both non ultimately periodic episturmian and non ultimately periodic balanced (see [6]) sequences over a 2-letter alphabet. Nevertheless, when the alphabet has 3 or more letters, the two notions no longer coincide. In particular, episturmian sequences are generally unbalanced over a k-letter alphabet, for $k \geq 3$. Thus, a natural question is to ask which sequences are both episturmian and balanced. We answer this question here.

We show that there are exactly three different kinds of balanced episturmian sequences, and among them, only one has different frequencies of letters. Moreover, this characterization gives a proof of the Fraenkel's conjecture [11, 10, 22, 21] for the special case of episturmian sequences. This conjecture was first introduced in number theory and has remained unsolved for more than 30 years. It states that for a fixed $k > 2$, there is only one way to cover $\mathbb{Z}$ by k Beatty sequences with pairwise distinct frequencies. The problem can be translated to combinatorics on words: for a k-letter alphabet, there is only one balanced sequence (up to letter permutation) that has different letter frequencies, which is called Fraenkel's sequence and is denoted by $(Fr_k)^\omega$ where $Fr_k = Fr_{k-1}kFr_{k-1}$, with $Fr_3 = 1213121$. The conjecture is verified for $k = 3, 4, 5, 6$ according to the work of Altman, Gaujal, Hordijk and Tijdeman [1, 12, 14]. The case $k = 7$ was recently settled by Barát and Varjú [4]. Many cases have been proved by Simpson [20]. Graham and O'Bryant [13] have generalized the conjecture to exact k-fold coverings and proved special cases of the generalized conjecture.

One interesting corollary of our main result is that Arnoux-Rauzy sequences [3] are never balanced episturmian sequences, since every balanced sequence is ultimately periodic.

In this paper, we first recall basic definitions and notations for combinatorics on words, as well as some useful results about episturmian sequences. Then, we show that balanced standard episturmian sequences over 3 or more letters are described by one of the following three directive sequences, up to letter permutations:

a) $\Delta(s) = 1^n 23 \ldots (k-1)k^\omega$, with $n \geq 1$;

b) $\Delta(s) = 12 \ldots (k-1)1k(k+1) \ldots (k+\ell-1)(k+\ell)^\omega$, with $\ell \geq 1$;

c) $\Delta(s) = 123 \ldots k1^\omega$,

where $k \geq 3$. As a result, since episturmian sequences have the same language as standard episturmian sequences, we prove a similar characterization for non-standard episturmian sequences. Finally, considering frequencies of letters in the balanced episturmian sequences, we prove Fraenkel's conjecture for the class of episturmian sequences.

2 Preliminaries

Let $\mathcal{A}$ denote a finite alphabet. A *finite word* w is an element of the free monoid $\mathcal{A}^*$. The i-th letter of w is denoted w_i. If $w = w_1 w_2 \ldots w_n$, with $w_i \in \mathcal{A}$, the length of w is n and we write $|w| = n$. By convention, the empty word is denoted ε and its length is 0. We define the set of non empty finite words as $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$, and $\mathcal{A}^\omega$ denotes the set of right infinite words over the alphabet $\mathcal{A}$, also called *sequences* for short. Then $\mathcal{A}^\infty = \mathcal{A}^* \cup \mathcal{A}^\omega$ is the set of finite and right infinite words over $\mathcal{A}$.

A word $w \in \mathcal{A}^\infty$ is *balanced* if for all factors u and v of w having the same length, one has $||u|_a - |v|_a| \leq 1$ for every $a \in \mathcal{A}$. A word $w \in \mathcal{A}^\infty$ is *ultimately periodic* of period $n \in \mathbb{N}$ if $w_i = w_{i+n} \quad \forall i \geq \ell$ and $\ell \in \mathbb{N}$. If $\ell = 1$, then w is *purely periodic*.

The number of occurences of the letter $a \in \mathcal{A}$ in w is denoted $|w|_a$. A word w is *a-free* if $|w|_a = 0$. For a finite word w, the frequency of the letter a is defined by $f_a(w) = |w|_a/|w|$. Note that we will compute the frequencies only for ultimately periodic and balanced sequences. Thus, the frequencies always exist (see [1]). The *reversal* of the finite word $w = w_1 w_2 \ldots w_n$ is $\widetilde{w} = w_n w_{n-1} \ldots w_1$ and if $\widetilde{w} = w$, then w is said to be a *palindrome*. A finite word f is a *factor* of $w \in \mathcal{A}^\infty$ if $w = pfs$ for some $p \in \mathcal{A}^*, s \in \mathcal{A}^\infty$. If $p = \varepsilon$ (resp. $s = \varepsilon$), f is called a *prefix* (resp. a *suffix*) of w. If $u = as$, $a \in \mathcal{A}$ and $s \in \mathcal{A}^\infty$, then, $a^{-1}u = s$. The *palindromic right closure* of $w \in \mathcal{A}^*$ is the shortest palindrome $u = w^{(+)}$ with w as prefix.

The set of factors of $s \in \mathcal{A}^\omega$ is denoted $F(s)$ and $F_n(s) = F(s) \cap \mathcal{A}^n$ is the set of all factors of s of length $n \in \mathbb{N}$. A factor f of s is *right* (resp. *left*) *special* in s if there exist $a, b \in \mathcal{A}$, $a \neq b$, such that $fa, fb \in F(s)$ (resp. $af, bf \in F(s)$). The *alphabet* of s is $\mathrm{Alph}(s) = F(s) \cap \mathcal{A}$ and $\mathrm{Ult}(s)$ is the set of letters occuring infinitely often in s.

Recall the definition of standard episturmian sequences introduced by Droubay, Justin and Pirillo:

Definition 2.1. ([9]) An infinite sequence s is *standard episturmian* if it satisfies one of the following equivalent conditions.

i) For any prefix u of s, $u^{(+)}$ is also a prefix of s.

ii) Every leftmost occurence of a palindrome in s is a central factor of a palindrome prefix of s.

iii) There exists an infinite sequence $u_1 = \varepsilon, u_2, u_3, \ldots$ of palindromes and an infinite sequence $\Delta(s) = x_1 x_2 \ldots.$, $x_i \in \mathcal{A}$, such that each of the words u_n defined by $u_{n+1} = (u_n x_n)^{(+)}$, $n \geq 1$, with $u_1 = \varepsilon$, is a prefix of s.

Notice that the proof of the equivalence of the conditions in Definition 2.1 is also given in [9].

Definition 2.2. ([9]) An infinite word t is *episturmian* if $F(t) = F(s)$ for some standard episturmian sequence s.

Notation 2.3. ([15]) Let $x = x_1 x_2 \ldots x_n$, $x_i \in \mathcal{A}$, and $u_1 = \varepsilon, \ldots, u_{n+1} = (u_n x_n)^{(+)}$, be the palindromic prefixes of u_{n+1}. Then $\mathrm{Pal}(x)$ denotes the word u_{n+1}.

In Definition 2.1, the word $\Delta(s)$ is called the *directive sequence* of the standard episturmian sequence s and we write $s = \mathrm{Pal}(\Delta(s))$.

Recall from [15] a useful property of the operator Pal. It will be used in almost all of our proofs in the next section.

Lemma 2.4. ([15]) *Let $x \in \mathcal{A}$. If w is x-free, then* $\mathrm{Pal}(wx) = \mathrm{Pal}(w)x\mathrm{Pal}(w)$. *If x occurs in w write $w = w'xw''$ with w'' x-free. Then, the longest palindromic prefix of* $\mathrm{Pal}(w)$ *which is followed by x in* $\mathrm{Pal}(w)$ *is* $\mathrm{Pal}(w')$; *whence easily* $\mathrm{Pal}(wx) = \mathrm{Pal}(w)\mathrm{Pal}(w')^{-1}\mathrm{Pal}(w)$.

Example 2.5. Let $w = \mathrm{Pal}(123) = 1213121$. Then, $\mathrm{Pal}(123 \cdot 4) = \mathrm{Pal}(123) \cdot 4 \cdot \mathrm{Pal}(123) = 121312141213121$ and $\mathrm{Pal}(1223 \cdot 2) = \mathrm{Pal}(1223) \cdot \mathrm{Pal}^{-1}(w') \cdot \mathrm{Pal}(1223) = 12121312121(121)^{-1}12121312121 = 1212131212121312121$, with $w = 1223$ and $w' = 12$.

The directive sequence allows to construct easily standard episturmian sequences.

Example 2.6. Over the alphabet $\mathcal{A} = \{1, 2, 3\}$, the Tribonacci sequence t (see [3]), a standard episturmian sequence, has the directive sequence $\Delta(t) = (123)^\omega$ and then, $u_1 = \varepsilon$, $u_2 = \underline{1}$, $u_3 = (1\underline{2})^{(+)} = 1\underline{2}1$, $u_4 = (121\underline{3})^{(+)} = 121\underline{3}121$, $\ldots$, $t = \underline{1}\underline{2}1\underline{3}121\underline{1}213121\underline{2}1312112131 21\ldots$

Remark. For clarity, we underline the letters of the directive sequence in the corresponding episturmian sequence.

Definition 2.7. A finite word $w \in \mathcal{A}^*$ is said to be *linear* if every letter $a \in \mathcal{A}$ occurs at most once.

Remark. If $1w$ is a linear word, then $\mathrm{Pal}(1w1) = (\mathrm{Pal}(1w))^2$. This formula will be used several times in the sequel.

Definition 2.8. Let w be a non-linear word over $\mathcal{A}$ and write $w = pw'$ with p the longest prefix of w which is linear. Then w'_1 is called the *first repeated letter* of w.

Examples 2.9. Let $u = 1432$, $v = 1212312$, and $w = 12321$. Then only u is linear, the first repeated letter of v is 1 and the one of w is 2.

A standard episturmian sequence $s \in \mathcal{A}^\omega$ (or any episturmian sequence with the same set of factors) is said to be $\mathcal{B}$-*strict* if $\mathrm{Ult}(\Delta(s)) = \mathrm{Alph}(s) = \mathcal{B} \subseteq \mathcal{A}$; that is every letter in $\mathcal{B} = \mathrm{Alph}(s)$ occurs infinitely many times in the directive sequence $\Delta(s)$. In particular, the $\mathcal{A}$-strict episturmian sequences correspond to the Arnoux-Rauzy sequences (see [3]).

3 Balanced episturmian sequences

In this section, we give a characterization of the balanced episturmian sequences over an alphabet with 3 or more letters. We first study balanced standard episturmian sequences and from that characterization, as standard episturmian sequences have the same language as episturmian sequences, we characterize more generally the balanced episturmian sequences.

Remark. In this paper, in order to show that a word is unbalanced, we will always give two factors f and f' of the same length, with $||f|_{f_1} - |f'|_{f_1}| \geq 2$, where the unbalance is over the first letter of f, namely f_1.

Let us start with some introductory examples.

Examples 3.1.

1) Let s be a standard episturmian sequence with the directive sequence $\Delta(s) = 1232\ldots$ Then,
$$s = \mathrm{Pal}(1232\ldots) = \underline{1}\underline{2}1\underline{3}121\underline{2}13121\cdots,$$
which contains the factors 212 and 131. Thus, s is unbalanced over the letter 2.

2) Let t be a standard episturmian sequence with the directive sequence $\Delta(t) = 12131\ldots$ Then
$$t = \mathrm{Pal}(12131\ldots) = \underline{1}\underline{2}1\underline{1}21\underline{3}1211211\underline{1}213121121\ldots,$$
which contains the factors 11211 and 21312. Thus, t is unbalanced over the letter 1.

3) Let u be a standard episturmian sequence with the directive sequence $\Delta(u) = 12341\ldots$ Then
$$u = \mathrm{Pal}(12341\ldots) = \underline{1}\underline{2}1\underline{3}121\underline{4}1213121\underline{1}21312141213121\ldots,$$
which is a balanced prefix.

It seems that the satisfiability of the balance condition depends on where the repeated letters occur. Proposition 3.3 characterizes directive sequences with the first repeated letter different from the first letter, while Proposition 3.4 characterizes the directive sequences $\Delta(s) = 11z$, with $z \in \mathcal{A}^\omega$.

Lemma 3.2. *Let $\Delta(s) = x\alpha^\ell y$ be the directive sequence of a balanced standard episturmian sequence s, with $\alpha \in \mathcal{A}$, $x \in \mathcal{A}^+$, $y \in \mathcal{A}^\omega$ and $\ell \geq 2$ maximum. If $x\alpha$ is linear, then* $\mathrm{Alph}(x) \cap \mathrm{Alph}(y) = \emptyset$.

Proof. Let suppose there exists a $\beta \in \mathrm{Alph}(x) \cap \mathrm{Alph}(y)$ such that $\Delta(s) = x'\beta x''\alpha^\ell y'\beta y''$, with $x'\beta x''y'$ linear. Let $p = \mathrm{Pal}(x'\beta x'')$. There are 3 cases:

a) If $x' \neq \varepsilon$, then
$$s = p(\underline{\alpha}p)^\ell \ldots (p\alpha)^\ell p\underline{\beta}p_1\ldots,$$
which contains the factors $\alpha p\alpha$ and $p\beta p_1$, $p_1 \neq \alpha$, $p_1 \neq \beta$. Thus, s is unbalanced.

b) If $x' = \varepsilon$ and $x'' \neq \varepsilon$, then
$$s = (p(\underline{\alpha}p)^\ell \ldots (p\alpha)^\ell p)^2\ldots,$$
which contains $\alpha p\alpha$ and pp_1p_2, $p_1 = \beta \neq \alpha$, $p_2 = x''_1 \neq \alpha$. Then s is unbalanced.

c) If $x' = \varepsilon$ and $x'' = \varepsilon$, then since s is over at least a 3-letter alphabet, there exists $\gamma \in \mathcal{A}$ such that at its first occurrence,
$$s = \underline{\beta}(\underline{\alpha}\beta)^\ell \ldots \beta\underline{\gamma}\beta\ldots,$$
which contains $\alpha\beta\alpha$ and $\beta\gamma\beta$, $\gamma \in \mathcal{A}$. Then, s is unbalanced. ∎

Proposition 3.3. *Let $\Delta(s)$ be the directive sequence of a balanced standard episturmian sequence s over a k-letter alphabet $\mathcal{A} = \{1, 2, \ldots, k\}$, $k \geq 3$. Let k be the first repeated letter of $\Delta(s)$. If $k \neq s_1$, then the directive sequence can be written as $\Delta(s) = 12\ldots(k-1)k^\omega$, up to letter permutation.*

Proof. Let $\Delta(s) = xkykz$ be the directive sequence of a balanced standard episturmian sequence s, with $x \in \mathcal{A}^+$, $y \in \mathcal{A}^*$, $z \in \mathcal{A}^\omega$, xky a linear word. Let $p = \mathrm{Pal}(x)$ and suppose $y \neq \varepsilon$. Then

$$s = p\underline{k}p\underline{y_1}pkp\ldots pkpy_1pkp\underline{k}p\ldots$$

which contains the factors kpk and py_1p_1, $p_1 \neq k$ and $y_1 \neq k$ (since xky is linear). Then s is unbalanced over k. Thus, $y = \varepsilon$ and so $\Delta(s) = xk^2z$. Assume $z \neq k^\omega$. We rewrite $\Delta(s) = xk^\ell z'$, with $z'_1 \neq k$, and $\ell \geq 2$. Since xk is linear, $z'_1 \notin \mathrm{Alph}(x)$ by Lemma 3.2. Thus xz'_1 is linear and

$$s = p(\underline{k}p)^\ell \underline{z}'_1 p_1 \ldots$$

which contains the factors kpk and pz'_1p_1, with $z'_1 \neq k$ and $p_1 = x_1 \neq k$. ∎

Proposition 3.4. *Let $\Delta(s)$ be the directive sequence of a balanced standard episturmian sequence s over a k-letter alphabet, $k \geq 3$. If $\Delta(s) = 1^\ell z$, with $z \in \mathcal{A}^\omega$, $z_1 \neq 1$ and $\ell \geq 2$, then $\Delta(s) = 1^\ell 23\ldots(k-1)k^\omega$, up to letter permutation.*

Proof. Let $\Delta(s) = 1^\ell z$ be the directive sequence of a balanced standard episturmian sequence s, with $z_1 \neq 1$, $\ell \geq 2$. Assume $|z|_1 > 0$. Then, $\Delta(s) = 1^\ell z' 1 z''$, with $z' \neq \varepsilon$ and $|z'|_1 = 0$. Since s is over at least a 3-letter alphabet, there exists at least one letter α in z' or z'' distinct from z'_1 and 1. At its first occurrence in s, it is preceded and followed by 1^ℓ. Then,

$$s = \underline{1}^\ell \underline{z}'_1 1^\ell \ldots 1^\ell z'_1 1^\ell \underline{1} z'_1 \ldots,$$

which contains the factors $z'_1 1^{\ell+1} z'_1$ and $1^\ell \alpha 1^2$, hence $|z|_1 = 0$. Since the alphabet is finite, there is at least one letter distinct from 1 which occurs twice in z. Let us consider the first repeated one in z, namely γ. Then, $\Delta(s) = 1^\ell u\gamma v\gamma w$, with $u\gamma v$ linear. Assume $v \neq \varepsilon$ and let $p = \mathrm{Pal}(1^\ell u)$. Then,

$$s = p\underline{\gamma}p\underline{v}_1 p\gamma p \ldots p\gamma p\underline{\gamma}\ldots$$

which contains the factors $\gamma p\gamma$, pv_1p_1, $v_1 \neq \gamma$ and $p_1 = 1$. It follows that $v = \varepsilon$. Let us now consider $\Delta(s) = 1^\ell u\gamma^2 w$, which we rewrite as $\Delta(s) = 1^\ell u\gamma^m w'$, u linear, $m \geq 2$ and assume $w'_1 \neq \gamma$. Then,

$$s = p(\underline{\gamma}p)^m \underline{w}'_1 p_1 \ldots$$

which contains the factors $\gamma p\gamma$ and pw'_1p_1. Hence, $w' = \gamma^\omega$ and the conclusion follows. ∎

Propositions 3.6 and 3.9 characterize the directive sequences $\Delta(s) = 1y1z$, with $y \neq \varepsilon$ and $1y$ linear. A technical lemma is required first:

Lemma 3.5. *Let $\Delta(s) = 1y1z$ be the directive sequence of a balanced standard episturmian sequence s, with $1y \in \mathcal{A}^*$ a linear word, $y \neq \varepsilon$, $z \in \mathcal{A}^\omega$. Then,* $\mathrm{Alph}(y) \cap \mathrm{Alph}(z) = \emptyset$.

Proof. Assume $\mathrm{Alph}(y) \cap \mathrm{Alph}(z) \neq \emptyset$ and let $\alpha \in \mathcal{A}$ be such that $\Delta(s) = 1y'\alpha y''1z'\alpha z''$, with $y'\alpha y''1z'$ a linear word. Let $p = \mathrm{Pal}(1y')$. Then, there are 2 cases:

a) If $y' \neq \varepsilon$, then using the equality $\mathrm{Pal}(1y'\alpha y'1) = (\mathrm{Pal}(1y'\alpha y'))^2$, it follows that

$$s = (p\underline{\alpha}p \dots p\alpha p)^2 \dots (p\alpha p \dots p\alpha p)^2 \underline{\alpha} \dots,$$

which contains the factors $\alpha p \alpha$ and $pp_1p_2 = p1y'_1$, $1 \neq \alpha$ and $y'_1 \neq \alpha$.

b) If $y' = \varepsilon$, then since s is over at least a 3-letter alphabet, there exists a letter $\beta \in \mathcal{A} \setminus \{1, \alpha\}$ in s and at its first occurrence, it is preceded and followed by 1. Then

$$s = (1\underline{\alpha}1 \dots 1\alpha 1)^2 \dots (1\alpha 1 \dots 1\alpha 1)^2 \underline{\alpha} \dots$$

and contains the factors $\alpha 1 \alpha$ and $1\beta 1$.

The conclusion follows. ■

In the following proposition, we describe how the letters different from 1 can be repeated.

Proposition 3.6. *Let $\Delta(s) = 1y1z$ be the directive sequence of a balanced standard episturmian sequence s over an alphabet with 3 or more letters. Suppose $|1y1z|_1 = 2$ and $y \neq \varepsilon$ a linear word. Thus, $\Delta(s) = 12 \dots (k-1)1k \dots (k+\ell-1)(k+\ell)^\omega$ (up to letter permutation), where $k \geq 3$.*

Proof. Let us consider $\Delta(s) = 1y1z$ with $|1y1z|_1 = 2$ and $\alpha \in \mathcal{A}$, the first repeated letter distinct from 1, that is the first repeated one in yz. Then, there are 2 cases to consider:

a) $\Delta(s) = 1y'\alpha y''1z'\alpha z''$, with $|y'y''1z'|_\alpha = 0$. Impossible, by Lemma 3.5.

b) $\Delta(s) = 1y'1z'\alpha z''\alpha z'''$, with $|y'1z'z''|_\alpha = |z'\alpha z''\alpha z'''|_1 = 0$. Assume $z'' \neq \varepsilon$ and let $p = \mathrm{Pal}(1y'1z')$. Then

$$s = p\underline{\alpha}p\underline{z}''_1 p\alpha p \dots p\alpha p z''_1 p\alpha p\underline{\alpha} \dots$$

with factors $\alpha p \alpha$ and pz''_1p_1. We conclude that $z'' = \varepsilon$.

The only possibility is $\Delta(s) = 1y'1z'\alpha^2 z'''$, which we rewrite as $\Delta(s) = 1y'1z'\alpha^\ell z''''$, $\ell \geq 2$, $z''''_1 \neq \alpha$. Let $p = \mathrm{Pal}(1y'1z')$. Then,

$$s = p(\underline{\alpha}p)^\ell \underline{z}''''_1 p_1 \dots$$

with factors $\alpha p \alpha$ and $pz''''_1 p_1$, where $z''''_1 \neq \alpha$, $p_1 = 1 \neq \alpha$. It follows that $z'''' = \alpha^\omega$, then $z''' = \alpha^\omega$, and finally, $\Delta(s) = 12 \dots k1(k+1) \dots (k+\ell)\alpha^\omega = 12 \dots k1(k+1) \dots (k+\ell)(k+\ell+1)^\omega$. ■

Example 3.7. Let s be a standard episturmian sequence with directive sequence $\Delta(s) = 12321\ldots$ Then,

$$s = \underline{12}1\underline{3}121\underline{2}13121\underline{1}213121213121\ldots$$

which contains the factors 212 and 131. Thus, s is unbalanced over the letter 2.

Example 3.8. Let t be a standard episturmian sequence with directive sequence $\Delta(t) = 12312\ldots$ Then,

$$t = \underline{12}1\underline{3}121\underline{1}213121\underline{2}13121\ldots$$

which contains the factors 212 and 131. Thus, t is unbalanced over the letter 2.

Proposition 3.9. *Let $\Delta(s) = 1y1z$ be the directive sequence of a balanced standard episturmian sequence s over a k-letter alphabet, $k \geq 3$, with $1y$ a linear word, $y \neq \varepsilon$ and $z \in \mathcal{A}^\omega$. If $z_1 \neq 1$, then $|z|_1 = 0$. If $z_1 = 1$, then $z = 1^\omega$ and $\Delta(s) = 123\ldots k(1)^\omega$ follows.*

Proof. Let us suppose $|z|_1 \geq 1$ (there is a third 1 in the directive sequence): $\Delta(s) = 1y1z'1z''$, with $|z'|_1 = 0$. Assume $z'_1 \neq \varepsilon$. Let $p = \mathrm{Pal}(1y)$. As $\mathrm{Pal}(1y1) = (\mathrm{Pal}(1y))^2 = p^2$, then

$$s = p^2\underline{z'_1}p^2\ldots p^2 z'_1 p^2 \underline{1}\ldots,$$

which contains the factors $1p1$ (in $p^2 1$) and $1^{-1}pz'_1p_1p_2$, where $z'_1 \neq p_1 = 1$ and $z'_1 \neq p_2 = y_1$ (ensured by Lemma 3.5). Then, $z'_1 = \varepsilon$ and $\Delta(s) = 1y11z''$. Rewrite $\Delta(s) = 1y1^\ell z'''$, with $\ell \geq 2$ and $z'''_1 \neq 1$. Let $p = \mathrm{Pal}(1y)$. As $\mathrm{Pal}(1y1^\ell) = (\mathrm{Pal}(1y))^{\ell+1}$ if $1y$ is linear,

$$s = p^{\ell+1}\ldots p^{\ell+1}\underline{z'''_1}1y_1\ldots$$

which contains the factors $1p1$ (in p^3) and $1^{-1}pz'''_1 1y_1$, with $y_1 \neq 1$. Then, s is unbalanced. It follows that $z = 1^\omega$. ■

We have now considered every possibility of directive sequences for a balanced standard episturmian sequence. Theorem 3.10 summarizes the previous propositions.

Theorem 3.10. *Any balanced standard episturmian sequence s over an alphabet with 3 or more letters has a directive sequence, up to a letter permutation, in one of the three following families of sequences:*

a) $\Delta(s) = 1^n\left(\prod_{i=2}^{k-1} i\right)(k)^\omega = 1^n 23\ldots(k-1)(k)^\omega$, *with $n \geq 1$;*

b) $\Delta(s) = \left(\prod_{i=1}^{k-1} i\right)1\left(\prod_{i=k}^{k+\ell-1} i\right)(k+\ell)^\omega = 12\ldots(k-1)1k\ldots(k+\ell-1)(k+\ell)^\omega$, *with $\ell \geq 1$;*

c) $\Delta(s) = \left(\prod_{i=1}^{k} i\right)(1)^\omega = 123\ldots k(1)^\omega$, *where* $k \geq 3$.

Proof. Proposition 3.3 implies a) for $n = 1$ and Proposition 3.4 implies a) for $n \geq 2$, while b) (resp. c)) follows from Proposition 3.6 (resp. Proposition 3.9). ■

Remark. Notice that all the directive sequences given in Theorem 3.10 yield balanced standard episturmian sequences. The sequences given in *a*) can be written as:

$$(1^n21^n31^n21^n41^n21^n31^n21^n \dots 1^n21^n31^n21^n41^n21^n31^n21^nk)^\omega.$$

This sequence is balanced over the letter 1, since the projection of s is $(1^n\alpha)^\omega$ which is balanced, and as the distance between two occurrences of the same letter is always the same, it follows from Hubert [14] that it is a balanced sequence. That is, the letter α is periodically replaced by $\mathrm{Pal}(23\dots(k-1))k$.

The sequences given in *b*) can be written as $Fr_{k-1}(Fr_{k-1}A)^\omega$, with A periodically replaced by $\mathrm{Pal}(k(k+1)\dots(k+\ell-1))(k+\ell)$. So, if the letter $i \in \{k, k+1, \dots, k+\ell\}$, then from Hubert [14] the sequence is balanced over this letter. The sequence is balanced over the letter $i \in \{1, 2, \dots, k-1\}$, since it appears in the Fraenkel word $(Fr_{k-1}AFr_{k-1})^\omega$.

Finally, the sequences given in *c*) are known to be balanced, from the Fraenkel's conjecture.

Recall the following result:

Theorem 3.11. ([9], Theorem 3) *A standard episturmian sequence s is ultimately periodic if and only if its directive sequence $\Delta(s)$ has the form $w\alpha^\omega$, $w \in \mathcal{A}^*$, $\alpha \in \mathcal{A}$.*

Then, a standard episturmian sequence cannot be both periodic and $\mathcal{A}$-*strict*. The next results follow easily.

Corollary 3.12. *Every balanced standard episturmian sequence on* 3 *or more letters is ultimately periodic.*

Proof. It is a direct consequence of Theorem 3.10. ■

Corollary 3.13. *None of the Arnoux-Rauzy sequences ($\mathcal{A}$-strict episturmian sequences) are balanced.*

Proof. It follows from the property that an $\mathcal{A}$-strict episturmian sequence cannot be ultimately periodic. ■

Remark. In [8], the authors have proved that one can construct an Arnoux-Rauzy sequence which is not c-balanced for any c. A sequence $s \in \mathcal{A}^\infty$ is *c-balanced* if for all factors u and v of s having the same length, one has $||u|_a - |v|_a| \leq c$ for every $a \in \mathcal{A}$

Corollary 3.14. *Any balanced standard episturmian sequence s, over an alphabet with more than* 2 *letters, is in one of the following families, up to letter permutation:*

a) $s = p(k-1)p\,(kp(k-1)p)^\omega$, *with* $p = \mathrm{Pal}(1^n2\dots(k-2))$;

b) $s = p(k+\ell-1)p\,[(k+\ell)p(k+\ell-1)p]^\omega$, *with* $p = \mathrm{Pal}(123\ldots(k-1)1k\ldots(k+\ell-2))$*;*

c) $s = [\mathrm{Pal}(123\ldots k)]^\omega$,

where $k \geq 3$.

Proof. It follows from the computation of $\mathrm{Pal}(\Delta(s))$, with $\Delta(s)$ given by one of the directive sequences of Theorem 3.10. ■

Proposition 3.15. *Every balanced standard episturmian sequence s, over an alphabet with 3 or more letters, with different frequencies for every letter can be written as in Corollary 3.14 c).*

Proof. In Corollary 3.14 a), the letters k and $(k-1)$ appear once in the period. Thus, the frequencies of k and $(k-1)$ in s are equal. In b), the same argument over the letters $(k+\ell)$ and $(k+\ell-1)$ holds. In c), by direct inspection, we find that the period of s has length $|\mathrm{Pal}(123\ldots k)| = 2^k - 1$, and the frequency of the letter i is $2^{k-i}/(2^k-1)$. Thus, the frequencies of two distinct letters are different. ■

As for every episturmian sequence t one could find a standard episturmian sequence s such that $F(s) = F(t)$, the result of Proposition 3.15 can be extended to any balanced episturmian sequence. Then, we get the following general result, which is a proof of Fraenkel's conjecture for episturmian sequences.

Theorem 3.16. *Let s be a balanced episturmian sequence over a k-letter alphabet $\mathcal{A} = \{1, 2, \ldots, k\}$, $k \geq 3$, with $f_i(s) \neq f_j(s)$, $\forall i \neq j$. In other words, every letter frequency is different. Then, up to letter permutation, $s = [\mathrm{Pal}(123\ldots k)]^\omega$.*

Example 3.17. *For* $k = 3, 4, 5$, *we obtain respectively* $s = (1213121)^\omega$, $t = (121312141213121)^\omega$ *and* $u = (1213121412131215121312141213121)^\omega$.

4 Concluding remarks

In order to extend our work on balanced sequences, we could investigate two directions.

The first one is in relation to billiard sequences. In [23], the second author surveys balanced sequences and proves that billiard sequences over a k-letter alphabet are $(k-1)$-balanced. That is, let X be a billiard sequence over a k-letter alphabet. Then

$$\forall i \in \mathcal{A}, \forall n \in \mathbb{N}, \forall w, w' \in F_n(X) \text{ we have } ||w|_i - |w'|_i| \leq k-1.$$

Furthermore, we notice that Fraenkel's sequences are periodic billiard sequences. Therefore, in the spirit of our paper, it would be interesting to study the billiard sequences [2, 7] which are balanced. This class contains at least Fraenkel's sequences, and perhaps some other interesting sequences.

The second direction is directly related to the original form of Fraenkel's conjecture. To prove this conjecture, it will be useful to have the property that balanced sequences over an alphabet with more than 2 letters and with pairwise distinct frequencies of letters, are given by directive sequences. Combining our results with this conjecture would give a proof of Fraenkel's conjecture.

5 Acknowledgements

We would like to thank Amy Glen who read carefully a previous version of this paper.

References

[1] E. Altman, B. Gaujal, A. Hordijk, *Balanced sequences and optimal routing*, J. ACM 47 (2000), no. 4, 752–775.

[2] P. Arnoux, C. Mauduit, I. Shiokawa, J.-I. Tamura, *Complexity of sequences defined by billiard in the cube*, Bull. Soc. Math. France 122 (1994), no. 1, 1–12.

[3] P. Arnoux, G. Rauzy, *Représentation géométrique de suites de complexité* $2n+1$, Bull. Soc. Math. France 119 (1991), no. 2, 199–215.

[4] J. Barát, P.P. Varjú, *Partitioning the positive integers into seven Beatty sequences*, Indag. Math. N.S. 12 (2003) 149–161.

[5] J. Berstel, *Recent results on extensions of Sturmian words*, International Conference on Geometric and Combinatorial Methods in Group Theory and Semigroup Theory (Lincoln, NE, 2000). Internat. J. Algebra Comput. 12 (2002), no. 1–2, 371–385.

[6] J. Berstel, P. Séébold, *Sturmian Words* in: Lothaire, Algebraic Combinatorics on Words, Cambridge University Press, 2002.

[7] J.-P. Borel, C. Reutenauer, *Palindromic factors of billiard words*, Theoret. Comput. Sci. 340 (2005), no. 2, 334–348.

[8] J. Cassaigne, S. Ferenczi, L. Q. Zamboni, *Imbalances in Arnoux-Rauzy sequences*, Ann. Inst. Fourier (Grenoble) 50 (2000), no. 4, 1265–1276.

[9] X. Droubay, J. Justin, G. Pirillo, *Episturmian words and some constructions of de Luca and Rauzy*, Theoret. Comput. Sci. 255 (2001) 539–553.

[10] A. Fraenkel, *The bracket function and complementary sets of integers*, Canad. J. Math. 21 (1969), 6–27.

[11] A. Fraenkel, *Complementing and exactly covering sequences*, J. Combinatorial Theory Ser. A 14 (1973), 8–20.

[12] R. L. Graham, *Covering the positive integers by disjoint sets of the form* $\{[n\alpha+\beta] : n = 1, 2, \ldots\}$, J. Combinatorial Theory Ser. A 15 (1973), 354–358.

[13] R.L. Graham, K. O'Bryant, *A discrete Fourier kernel and Fraenkel's tiling conjecture*, Acta Arith. 118 (2005), no. 3, 283–304

[14] P. Hubert, *Suites équilibrées*, Theoret. Comput. Sci. 242 (2000), no. 1–2, 91–108.

[15] J. Justin, *Episturmian morphisms and a Galois theorem on continued fractions*, Theor. Inform. Appl. 39 (2005), 207–215.

[16] J. Justin, L. Vuillon, *Return words in Sturmian and episturmian sequences*, Theor. Inform. Appl. 34 (2000), no. 5, 343–356.

[17] A. de Luca, *Sturmian words: structure, combinatorics and their arithmetics*, Theoret. Comput. Sci. 183 (1997) 45–82.

[18] M. Morse, G.A. Hedlund, *Symbolic dynamics II - Sturmian trajectories*, Amer. J. Math. 62 (1940), 1–42.

[19] G. Rauzy, *Mots infinis en arithmétique,* in: M. Nivat, D. Perrin (Eds.), Automata on Infinite Words, Lecture Notes in Computer Science, Vol 192 (1985), 167–171.

[20] R.J. Simpson, *Disjoint covering systems of rational Beatty sequences*, Discrete Math. 92 (1991), 361–369.

[21] R. Tijdeman, *Exact covers of balanced sequences and Fraenkel's conjecture*, Algebraic number theory and Diophantine analysis (Graz, 1998), 467–483, de Gruyter, Berlin, 2000.

[22] R. Tijdeman, *Fraenkel's conjecture for six sequences*, Discrete Math. 222 (2000), no. 1–3, 223–234.

[23] L. Vuillon, *Balanced words*, Bull. Belg. Math. Soc. Simon Stevin 10 (2003), suppl., 787–805.