Guest Editor
Peter Neumann

The Complexity of Songs

DONALD E. KNUTH

Every day brings new evidence that the concepts of computer science are applicable to areas of life which have little or nothing to do with computers. The purpose of this survey paper is to demonstrate that important aspects of *popular songs* are best understood in terms of modern complexity theory.

It is known [3] that almost all songs of length n require a text of length $\sim n$. But this puts a considerable space requirement on one's memory if many songs are to be learned; hence, our ancient ancestors invented the concept of a refrain [14]. When the song has a refrain, its space complexity can be reduced to cn, where $c < 1$ as shown by the following lemma.

LEMMA 1.
Let S be a song containing m verses of length V and a refrain of length R where the refrain is to be sung first, last, and between adjacent verses. Then, the space complexity of S is $(V/(V + R)) n + O(1)$ for fixed V and R as $m \to \infty$.

PROOF.
The length of S when sung is

$$n = R + (V + R)m \tag{1}$$

while its space complexity is

$$c = R + Vm. \tag{2}$$

By the Distributive Law and the Commutative Law [4], we have

$$c = n - (V + R)m + mV$$
$$= n - Vm - Rm + Vm \tag{3}$$
$$= n - Rm.$$

The lemma follows. □

[It is possible to generalize this lemma to the case of verses of differing lengths $V_1, V_2, \ldots, V_m$, provided that the sequence $\langle V_k \rangle$ satisfies a certain smoothness condition. Details will appear in a future paper.]

A significant improvement on Lemma 1 was discovered in medieval European Jewish communities where an anonymous composer was able to reduce the complexity to $O(\sqrt{n})$. His song "Ehad Mi Yode'a" or "Who Knows One?" is still traditionally sung near the end of the Passover ritual, reportedly in order to keep the children awake [6]. It consists of a refrain and 13 verses $v_1, \ldots, v_{13}$, where v_k is followed by $v_{k-1} \ldots v_2 v_1$ before the refrain is repeated; hence m verses of text lead to $\frac{1}{2}m^2 + O(m)$ verses of singing. A similar song called "Green Grow the Rushes O" or "The Dilly Song" is often sung in western Britain at Easter time [1], but it has only twelve verses (see [1]), where Breton, Flemish, German, Greek, Medieval Latin, Moravian, and Scottish versions are cited.

The coefficient of $\sqrt{n}$ was further improved by a Scottish farmer named O. MacDonald, whose construction[1] appears in Lemma 2.

[1] Actually MacDonald's priority has been disputed by some scholars; Peter Kennedy ([8], p. 676) claims that "I Bought Myself a Cock" and similar farmyard songs are actually much older.

Lemma 2.
Given positive integers α and λ, there exists a song whose complexity is $(20 + \lambda + \alpha)\sqrt{n}/(30 + 2\lambda) + O(1)$.

Proof.
Consider the following schema [9].

$$V_O = \text{'Old MacDonald had a farm, ' } R_1$$

$$R_1 = \text{'Ee-igh, '}^2 \text{ 'oh! '}$$

$$R_2(x) = V_O \text{ 'And on this farm he had}$$

$$\text{some' } x\text{', ' } R_1 \text{ 'With a'}$$

$$U_1(x, x') = x\text{', ' } x' \text{ ' here and a ' } x\text{', ' } x' \text{ ' there; '}$$

$$U_2(x, y) = x\text{'here a ' } y, \text{ ' '}$$

$$U_3(x, x') = U_1(x, x') U_2(\epsilon, x) U_2(\text{'t'}, x')$$

$$U_2(\text{'everyw'}, x\text{', ' } x')$$

$$V_k = U_3(W_k, W'_k)V_{k-1} \quad \text{for} \quad k \geq 1$$

where

$$W_1 = \text{'chick'}, \ W_2 = \text{'quack'},$$

$$W_3 = \text{'gobble'}, \ W_4 = \text{'oink'}, \tag{5}$$

$$W_5 = \text{'moo'}, \ W_6 = \text{'hee'},$$

and

$$W'_k = W_k \quad \text{for} \quad k \neq 6; \ W'_6 = \text{'haw'}. \tag{6}$$

The song of order m is defined by

$$\mathcal{S}_0 = \varepsilon \tag{7}$$

$$\mathcal{S}_m = R_2(W''_m)V_m\mathcal{S}_{m-1} \quad \text{for} \quad m \geq 1,$$

where

$$W''_1 = \text{'chicks'}, \ W''_2 = \text{'ducks'},$$

$$W''_3 = \text{'turkeys'}, \ W''_4 = \text{'pigs'}, \tag{8}$$

$$W''_5 = \text{'cows'}, \ W''_6 = \text{'donkeys'}.$$

The length of $\mathcal{S}(m)$ is

$$n = 30m^2 + 153m$$

$$+ 4(m\ell_1 + (m - 1)\ell_2 + \ldots + \ell_m)$$

$$+ (a_1 + \ldots + a_m) \tag{9}$$

while the length of the corresponding schema is

$$c = 20m + 211 + (\ell_1 + \ldots + \ell_m)$$

$$+ (a_1 + \ldots + a_m). \tag{10}$$

Here $\ell_k = |W_k| + |W'_k|$ and $a_k = |W''_k|$, where $|x|$ denotes the length of string x. The result follows at once, if we assume that $\ell_k = \lambda$ and $a_k = \alpha$ for all large k. □

Note that the coefficient $(20 + \lambda + \alpha)/\sqrt{30 + 2\lambda}$ assumes its minimum value at

$$\lambda = \max(1, \alpha{-}10) \tag{11}$$

when α is fixed. Therefore if MacDonald's farm animals ultimately have long names they should make slightly shorter noises.

Similar results were achieved by a French Canadian ornithologist, who named his song schema "Alouette" [2, 15]; and at about the same time by a Tyrolean butcher whose schema [5] is popularly called "Ist das nicht ein Schnitzelbank?" Several other cumulative songs have been collected by Peter Kennedy [8], including "The Mallard" with 17 verses and "The Barley Mow" with 18. More recent compositions, like "There's a Hole in the Bottom of the Sea" and "I Know an Old Lady Who Swallowed a Fly" unfortunately have comparatively large coefficients.

A fundamental improvement was claimed in England in 1824, when the true love of U. Jack gave to him a total of 12 ladies dancing, 22 lords a-leaping, 30 drummers drumming, 36 pipers piping, 40 maids a-milking, 42 swans a-swimming, 42 geese a-laying, 40 golden rings, 36 collie birds, 30 french hens, 22 turtle doves, and 12 partridges in pear trees during the twelve days of Christmas [11]. This amounted to $\frac{1}{6} m^3 + \frac{1}{2} m^2 + \frac{1}{3} m$ gifts in m days, so the complexity appeared to be $O(\sqrt[3]{n})$; however, it was soon pointed out [10] that his computation was based on n *gifts* rather than n units of singing. A complexity of order $\sqrt{n}/\log n$ was finally established (see [7]).

We have seen that the partridge in the pear tree gave an improvement of only $1/\sqrt{\log n}$; but the importance of this discovery should not be underestimated since it showed that the $n^{0.5}$ barrier could be broken. The next big breakthrough was in fact obtained by generalizing the partridge schema in a remarkably simple way. It was J. W. Blatz of Milwaukee, Wisconsin who first discovered a class of songs known as "m Bottles of Beer on the Wall"; her elegant construction[2] appears in the following proof of the first major result in the theory.

Theorem 1.
There exist songs of complexity $O(\log n)$.

Proof.
Consider the schema

$$V_k = T_kBW\text{', '}$$

$$T_kB\text{'; '}$$

$$\text{If one of those bottles should} \tag{12}$$

$$\text{happen to fall, '}$$

$$T_{k-1}B \ W\text{'.'}$$

where

$$B = \text{' bottles of beer'}, \tag{13}$$

$$W = \text{' on the wall'},$$

and where T_k is a translation of the integer k into Eng-

---

[2] Again Kennedy ([8], p. 631) claims priority for the English, in this case because of the song "I'll drink m if you'll drink $m + 1$." However, the English start at $m = 1$ and get no higher than $m = 9$, possibly because they actually drink the beer instead of allowing the bottles to fall.

lish. It requires only $O(m)$ space to define T_k for all $k < 10^m$ since we can define

$$T_{q \cdot 10^m + r} = T_q \text{ ' times 10 to the ' } T_m \text{ ' plus ' } T_r \quad (14)$$

for $1 \leq q \leq 9$ and $0 \leq r < 10^{m-1}$.

Therefore the songs S_k defined by

$$S_O = \varepsilon, \quad S_k = V_k S_{k-1} \quad \text{for} \quad k \geq 1 \quad (15)$$

have length $n \asymp k \log k$, but the schema which defines them has length $O(\log k)$; the result follows. □

Theorem 1 was the best result known until recently[3], perhaps because it satisfied all practical requirements for song generation with limited memory space. In fact, 99 bottles of beer usually seemed to be more than sufficient in most cases.

However, the advent of modern drugs has led to demands for still less memory, and the ultimate improvement of Theorem 1 has consequently just been announced:

THEOREM 2.
There exist arbitrarily long songs of complexity $O(1)$.

PROOF: (due to Casey and the Sunshine Band). Consider the songs S_k defined by (15), but with

$$V_k = \text{'That's the way,' } U \text{ 'I like it, ' } U \quad (16)$$

$$U = \text{'uh huh,' 'uh huh'}$$

for all k. □

It remains an open problem to study the complexity of nondeterministic songs.

---

[3] The chief rival for this honor was "This old man, he played m, he played knick-knack . . .".

REFERENCES
1. Rev. S. Baring-Gould, Rev. H. Fleetwood Sheppard, and F.W. Bussell. *Songs of the West* (London: Methuen, 1905), 23, 160–161.
2. Oscar Brand. *Singing Holidays* (New York: Alfred Knopf, 1957), 68–69.
3. G.J. Chaitin. "On the length of programs for computing finite binary sequences: Statistical considerations." *J. ACM* 16 (1969), 145–159.
4. G. Chrystal. *Algebra, an Elementary Textbook* (Edinburgh: Adam and Charles Black, 1886), Chapter 1.
5. A. Dörrer, *Tiroler Fasnacht* (Wien, 1949), 480 pp.
6. *Encyclopedia Judaica* (New York: Macmillan, 1971), v. 6 p. 503; *The Jewish Encyclopedia* (New York: Funk and Wagnalls, 1903); articles on Ehad Mi Yode'a.
7. U. Jack, "Logarithmic growth of verses," *Acta Perdix* 15 (1826), 1–65535.
8. Peter Kennedy, *Folksongs of Britain and Ireland* (New York: Schirmer, 1975), 824 pp.
9. Norman Lloyd, *The New Golden Song Book* (New York: Golden Press, 1955), 20–21.
10. N. Picker, "Once señores brincando al mismo tiempo," *Acta Perdix* 12 (1825), 1009.
11. ben shahn, a partridge in a pear tree (New York: the museum of modern art, 1949), 28 pp. (unnumbered).
12. Cecil J. Sharp, ed., *One Hundred English Folksongs* (Boston: Oliver Ditson, 1916), xlii.
13. Christopher J. Shaw, "that old favorite, Apiapt/a Christmastime algorithm," with illustrations by Gene Oltan, *Datamation* 10, 12 (December 1964), 48–49. Reprinted in Jack Moshman, ed., *Faith, Hope and Parity* (Washington, D.C.: Thompson, 1966), 48–51.
14. Gustav Thurau, *Beiträge zur Geschichte und Charakteristik des Refrains in der französischen Chanson* (Weimar: Felber, 1899), 47 pp.
15. Marcel Vigneras, ed., *Chansons de France* (Boston: D.C. Heath, 1941), 52 pp.