

articles

A Generative Textsetting Model

By John Halle and Fred Lerdahl

One aspect of musical practice that has received comparatively little attention in recent years is the system of intuitions operating when a composer assigns notes to words. In contrast to the extremely varied compositional techniques of Western music, composers' impulses are narrowly constrained by both musical and linguistic intuitions in textsetting. That this is the case can be seen by attempting to match the lines of poetry with their associated rhythmic settings:

(1a) Tell me not in mournful numbers.

—Longfellow, "A Psalm of Life," line 1

(1b) Through all the compass of the notes.


—Dryden, "A Song for St. Cecilia's Day," line 15

(1c) | ♪ ♪ ♪ ♪ | ♪ ♪ ♪ ♪ |

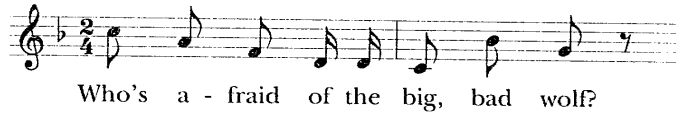
(1d) | 7 ♪ ♪ ♪ | ♪ ♪ ♪ ♪ | ♪

One need not have had much musical training to know to pair (1a) with (1c), and (1b) with (1d). While it may appear obvious that one's judgments are informed by intuitions in attempting to match "stressed" syllables with "strong" beats, the problem that confronts the theorist attempting to explain this system of intuitions is more complex than an initial formulation suggests.

Our approach to this problem makes a basic methodological simplification. As indicated by the absence of pitch notation in the above examples, we assume that on a local level, at least, the notion of the "strong beat" relevant to textsetting is predominantly a rhythmic and metrical phenomenon independent of pitch height. For example, a rising sequence of pitches might in principle seem a more natural setting for the rising intonational pattern between "bad" and "wolf" in the question:


Who's afraid of the big, bad wolf?

Yet most children know that "bad wolf" is set by a falling pair of pitches:



While the relationship of musical and phonological pitch contour is highly constrained in settings of tone languages such as Chinese,¹ as a rule the correspondence of musical and phonological pitch contours is a secondary consideration for textsetting in the idioms with which we shall concern ourselves. Much narrower constraints are imposed by the requirements, exemplified in (1), of assigning “strong” metrical positions to “stressed” syllables. Our notation will reflect this fact.

The notation in (1c) and (1d) indicates a series of attack points and durations without pitch. Rhythm, however, does not exist solely in the events themselves but is a structure that experienced listeners infer from particular sequences of events. Although the two settings indicate identical series of durations, they signify two fundamentally distinct metrical structures: (1c) represents a “weak-strong” pattern, while (1d) is “strong-weak.” It is the inferred structure of strong and weak events that listeners assign to patterns, rather than their acoustic organization, which interests us here. To capture the former, we shall follow Fred Lerdahl and Ray Jackendoff’s *A Generative Theory of Tonal Music* (hereafter *GTTM*)² in assigning grid representations beneath conventionally notated series of attacks and durations.

The “Metrical Well-Formedness and Preference” rule system outlined in *GTTM* assigns the following grid structures to the patterns in (1c) and (1d):

(2a) | |
 * * * * * * * *
 * * * * * * * *
 * * * * * * * *

(2b) | |
 * * * * * * * *
 * * * * * * * *
 * * * * * * * *

¹ Bell Yung, “The Relationship of Text and Tune in Chinese Opera,” in *Music, Language, Speech, and Brain*, ed. Johan Sundberg, Lennart Nord, and Rolf Carlson (London: Macmillan, 1991), 408–18.

² Fred Lerdahl and Ray Jackendoff, *A Generative Theory of Tonal Music* (Cambridge: MIT Press, 1983), 68–104.

We shall not discuss the system of rules, outlined in *GTTM*, that motivate a listener's assignment of a grid to durational sequences.³ It suffices to point out that intuitively based judgments are directly represented by (2a) and (2b), as opposed to the conventional notation in (1c) and (1d). There is no indication in (1c), for example, that the third eighth note of each measure is to be heard as stronger than the second and fourth.

A listener hears a musical surface as "having a beat" largely insofar as he can easily assign it a grid. Having done so, the listener will generally hear a single metrical level as most prominent. This level, defined in *GTTM* as the *tactus*, corresponds to what is informally spoken of as "the beat." It is in reference to this level that most activities carried out with musical accompaniment—such as dancing, jump-roping, and marching—are synchronized. A piece is heard as "in one," as opposed to "in two" or "in four," depending on where the *tactus* is located.

Given the perceptual prominence of the *tactus*, we suggest a notational refinement of the *GTTM* grid system by defining rhythmic levels in relation to it. The *tactus* will be considered level 0, or L(0), with levels above or below denoted by positive or negative integers. As shown in (3), smaller levels are assigned a numerically lower index, and larger levels are given a higher index:



Events that take place on multiple levels will be said to "occur on" or "be assigned to" the most prominent level on which they are situated. Thus, while the first and the second events in (3) both correspond to an L(-1) position, only the latter is assigned to level L(-1), for it is the largest level on which the second event occurs. The former is assigned to level L(1).

* * *

Having suggested a means for representing the hierarchy that a listener assigns to rhythmic events, we now turn to the corresponding problem of how to represent the linguistic hierarchy of strong and weak syllables embodied in the varying degrees of stress that a speaker assigns to words and phrases. In music, as pointed out above, metrically strong and weak

³ Ibid.

6 CURRENT MUSICOLOGY

beats are relative notions: an event may be strong in relation to a second event but weak in relation to a third. A similar situation obtains in language. While one tends to speak of a specific syllable of a word as being accented or strong, in practice most polysyllabic words or phrases are composed of two or more stressed syllables, some of which are more strongly accented than others. In the word "Ticonderoga," for example, the first syllable is weak compared to the fourth but strong relative to the third. As in the musical case, a grid representation accurately represents the stress hierarchy. Morris Halle and Jean-Roget Vergnaud present the following phonological grid for "Ticonderoga":⁴

					x	line (3)
					x	line (2)
	x				x	line (1)
	x	x			x	line (0)
Ti	con	de	ro	ga		

The process by which a speaker derives a grid from phonological and morphological input is a subject of considerable discussion in the field of generative phonology,⁵ and is therefore well beyond the scope of this paper. As with rhythmic settings, we are concerned not with how a grid is assigned but with the structure it represents. In the phonological grid, all syllables are assigned to a position on line(0) with unstressed syllables, realized in English by the reduced vowel sound "schwa," receiving an x only on this line. Stressed syllables receive x's above this level according to their degree of stress. Primary stressed syllables receive the highest column of x's on the grid.

Grid structures can be also be assigned, albeit less definitively, to larger linguistic units such as phrases, sentences and sequences of sentences. For example, a normal delivery of "Belgian farmers grow turnips" will manifest the following grid structure:⁶

							x	line (3)
							x	line (2)
	x						x	line (1)
	x	x	x	x	x	x	x	line (0)
Bel-gian	farm-ers	grow	tur-nips.					

⁴ Morris Halle and Jean-Roget Vergnaud, *An Essay on Stress* (Cambridge: MIT Press, 1987), 9.

⁵ See, for example, Alan Prince, "Relating to the Grid" *Linguistic Inquiry* 14 (1983): 19-100, and Mark Liberman, *The Intonational System of English* (Dissertation at MIT, 1975).

⁶ Bruce Hayes, "The Phonology of Rhythm in English," *Linguistic Inquiry* 15 (1984): 35.

Accented syllables are those receiving stress on line 2 or above on the phonological grid.

The observation that in setting texts one tends to assign accented syllables to strong beats is now precisely defined: syllables are either accented or unaccented according to the height of their associated column on the phonological grid; metrical positions are either strong or weak according to their location in the metrical grid. This principle ignores the assignment of unaccented syllables, which may appear in either strong or weak musical positions, as evidenced by the placement of “of” in (4a). The prohibition against the appearance of certain types of stressed syllables in weak positions is mirrored in similar restrictions operative in poetic meters.⁸

* * *

Our inquiry might now continue in several directions. One possibility is to test and refine a system of textsetting rules based on compositional practice. We shall not explore this approach here, since in creating vocal music, composers are often interested in exploring unusual textsetting possibilities. Options that contradict basic impulses may be preferred by composers precisely because they are violations and hence are striking and unexpected. Given our empirical interest in an unambiguous application of basic textsetting principles, we prefer to examine a musical context in which the simplest solutions are encouraged, rather than rejected as too obvious.

Group singing in church services, folk-song singalongs, jump-roping songs, work songs or marching chants offer controlled environments for studying basic textsetting intuitions. In such contexts, most of those present will know a tune and perhaps the first verse of the text, but often not the subsequent verses of the text, in which case they are listed beneath the first verse as if they were stanzas of poetry. Or, on occasion, new texts are shouted out by the leader immediately before the next verse is sung, with no specific indication as to how the music and text are to correspond. Quite frequently the “tune” that sets the first verse needs to be varied substantially from the original in order to accommodate successive verses. This essentially creative process can be left to the intuitions of even inexperienced singers, a group of whom can be relied on to produce settings that are sufficiently similar for the ensemble to sound together. The musi-

⁸ Morris Halle and Jay Keyser, *English Stress: Its Form, Its Growth, and its Role in Verse* (New York: Harper and Row, 1968), 169.

cal variants created for successive stanzas are suggested by the structures of the texts, which, in conjunction with the maintenance of the structure of the tune, generate a preferred setting. All those singing have internalized the basic principles involved; otherwise the degree of uniformity would not be apparent.

The Anglo-American sea chanty *The Drunken Sailor*⁹ is typical of the sort of material that might be performed in such contexts:

The Drunken Sailor

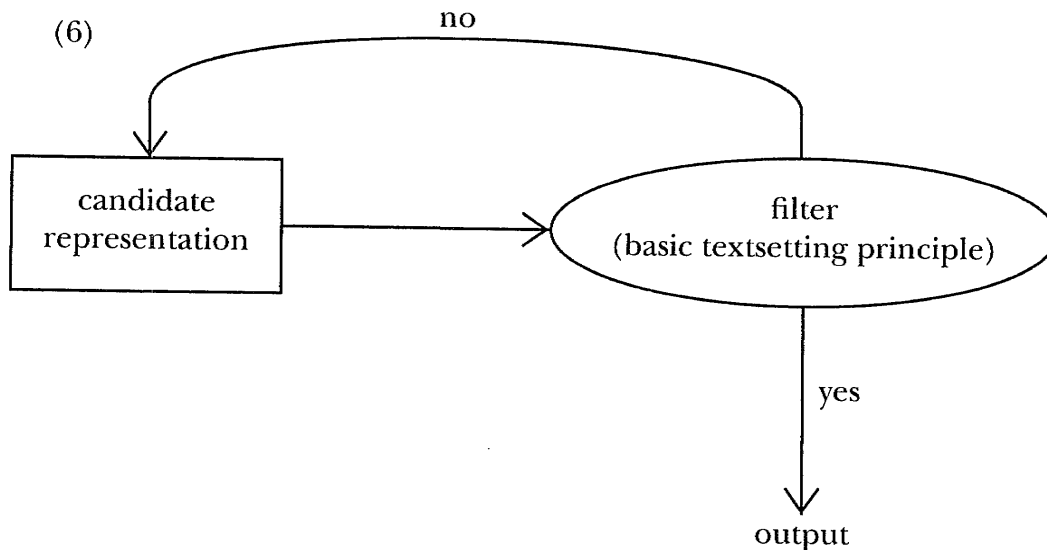
What shall we do with the drunk - en sail - or? What shall we do with the
drunk - en sail - or? What shall we do with the drunk - en sail - or
ear - ly in the morn - ing!

2. Put him in the guard room till he gets sober. (Three times.)
3. Keep him there an' make him bail her.
4. Trice him up in a runnin' bowline.
5. Tie him to the taffrail when she's yard-arm under.
6. Put him in the scuppers with a hose-pipe on him.
7. What shall we do with the Queen o' Sheba?
8. Keel-haul him 'til he's sober.
9. Give him a taste o' the bosun's rope-end.
10. Stick on his back a mustard plaster.
11. What'll we do with a Limejuice Skipper?
12. Soak him in oil till he sprouts a flipper.
13. Scrape the hair off his chest with hoop-iron razor.

A basic component of this style is that each syllable must be set by one note. Thus, each of the verses having fewer than ten syllables must be set to melodies having fewer notes than the initial statement, while those with more than ten syllables can be accommodated only by settings having more than ten. The rhythm of the melody as initially presented must be altered substantially in order to accommodate the texts of ensuing verses.

⁹ John Ashton, ed., *Real Sailor Songs* (London: Simpkin, Marshall, Hamilton and Kent, 1891).

As suggested in our previous discussion, however, only certain variants of the melody are judged as acceptable settings and are realized in performance. In a conventional textsetting environment like a sea-chanty, “accented syllables” (as defined above) must correspond with “strong beats” (as defined above). We shall refer to this procedure as “the basic textsetting principle.” An explanation of the process by which singers generate tunes might take the form of (6): a candidate setting is generated, and then must pass through a filter which incorporates the intuitions defined by the basic textsetting principle.



The filter passes through acceptable settings such as (7a) and (7b), but rejects unacceptable settings such as (7c).

The theoretical framework implicit in (6) is flawed in two respects. First, as a constraint on the output it is inefficient: a class of inputs is generated and subsequently fed into the filter, which churns through the possibilities, rejecting most of them until an acceptable candidate is generated. The candidate (7c) is only one member of a large class that is rejected by the filter because of the assignment of “make” and “bail” to weak rhythmic positions. More serious, however, is that the filter also admits (7b), which is clearly an unnatural setting. A more effective approach is to devise constraints on the input to the filter, in addition to constraints on the output. Indeed, if sufficiently rigid constraints can be set on the generation of settings, the filter as represented in (6) will be unnecessary.

The most significant of these constraints derives from the paradigmatic tune that initially sets the opening stanza. As we have observed, literal maintenance of the rhythms of the paradigm makes it impossible to accommodate successive stanzas having different syllable counts from the

(7a)

x					x			x	line (3)	
x					x			x	line (2)	
x					x			x	line (1)	
x	x	x	x		x	x		x	x	line (0)
Keep	him	there	and	make	him	bail	her.			
*	*	*	*	*	*	*	*	L(-1)		
*		*		*		*		L(0)		
*				*		*		L(1)		

(7b)

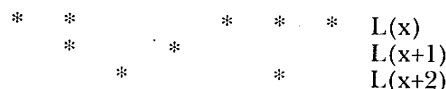
								x	line (3)	
								x	line (2)	
								x	line (1)	
	x	x	x		x	x		x	x	line (0)
Keep	him	there	and	make	him	bail	her.			
*	*	*	*	*	*	*	*	L(-2)		
*		*		*		*		L(-1)		
*				*		*		L(0)		
*						*		L(1)		

(7c)

								x	line (3)	
								x	line (2)	
								x	line (1)	
	x	x	x	x	x	x		x	line (0)	
Keep	him	there	and	make	him	bail	her.			
*	*	*	*	*	*	*	*	L(-2)		
*	*		*	*	*	*	*	L(-1)		
*			*		*		*	L(0)		
*				*		*		L(1)		

initial stanza. The question is what specifically must be maintained from the paradigm. We propose that in hearing *The Drunken Sailor*, one derives a metrical grid that generates a finite repertoire of rhythmic sequences to which all settings are subsequently made to conform. The principal characteristics of this grid are defined by what will be referred to as Metrical Well-Formedness Rules (hereafter MWFRs).

The first two of such rules are general conditions that apply to the geometry of metrical grid structures. MWFR 1, which states that a beat on any level is also a beat on all smaller levels, prohibits structures such as the following which, while perhaps signifying the disposition of voices in a contrapuntal texture, do not represent a metrical hierarchy:



MWFR 2, which requires equally spaced beats, assumes periodicity as a necessary condition for the perception of metrical structure. While acoustical events may not exhibit a perfect regularity, the beats to which they correspond must be understood as functionally equidistant.

MWFRs 3, 4, and 5 refer to specific characteristics of the paradigmatic grid for *The Drunken Sailor* settings. As remarked earlier, one level of the rhythmic hierarchy—the tactus level $L(0)$ —defines a “beat,” in reference to which physical activities tend to be choreographed. A basic characteristic of *The Drunken Sailor* is that each line is set by a sequence of durations containing four strong beats. The line is then repeated twice, maintaining intact the durational sequence and its associated four-beat grid. Subsequent settings of new stanzas, while departing significantly from the initial sequence, always maintain the framework of these four beats (as indicated in MWFR 3). The reader can confirm this by performing the song with all the verses and tapping at the intuitively most natural points.

The second constraint on *The Drunken Sailor* grids follows from the observation that triple rhythms (triplet-quarter or -eighth notes) are excluded from consideration as settings for these verses. While ternary subdivisions are of course possible in principle, we restrict ourselves to generating settings that contain exclusively binary subdivisions (as stated in MWFR 4): sixteenth, eighth, quarter, half notes—never triplets or dotted rhythms.

A final constraint on grids expresses the acoustical or physiological fact that the intonation of a syllable requires some minimal duration to be either understood by a listener or managed by the speaker (or singer). Hence MWFR 5 states that no event may be situated on a metrical level $L(x)$ where $x \geq 2$. This excludes settings that make use of units less than the sixteenth note in the above transcriptions. For convenience we restate MWFRs 1–5:

- MWFR 1.** A beat on any level is also a beat on all smaller levels.
- MWFR 2.** All levels consist of equally spaced beats
- MWFR 3.** All *Drunken Sailor* grids must contain exactly four $L(0)$ beats.
- MWFR 4.** In *Drunken Sailor* grids, beats on $L(x)$ are equally subdivided by one beat at $L(x-1)$.
- MWFR 5.** No event may be situated on a metrical level $L(x)$ where $x \geq 2$.

Finally, we state the requirement that, since each line of text is repeated twice, each repetition must be synchronized with the renewed onset of the four-beat grid. A setting that imposes a line boundary before the completion of the metrical grid is rejected as a violation of TWFR 3, which requires that each line of text occupy the full extent of the paradigmatic grid. One such case is the following:

Keel	haul	him	till	he's	so	-	ber./	Keel	haul	...	
	♪	♪	♪	♪	♪		♪	♪	♪	♪	
*	*	*	*	*	*	*	*	*	*	*	*
*	*	*	*	*	*	*	*	*	*	*	L(-2)
*	*	*	*	*	*	*	*	*	*	*	L(-1)
*	*	*	*	*	*	*	*	*	*	*	L(0)

This setting is a violation of TWFR 3. We restate all TWFRs in sequence:

TWFR 1. In all settings, each syllable is associated with a beat on some level of the grid.

TWFR 2. Each syllable occupies the entire time span up to, but not including, the beat corresponding to the onset of the successive syllable.

TWFR 3. Each line of text must occupy the full extent of the paradigmatic grid.

We now have at our disposal rules that constrain both acceptable rhythmic sequences and possible mappings between syllables and rhythms. Unlike the textsetting model invoked in (6), these constraints ensure a limited input. Next we propose an algorithm (9) that, in conjunction with well-formedness conditions, assigns a unique setting to texts.

(9) Textsetting Algorithm

Step 1. Assign all accented syllables to available L(0) beats from left to right.

Step 2. Assign syllables to L(-1) beats from left to right.

Step 3. Assign syllables to L(-2) beats from left to right.


Let us apply these three steps to the second stanza, “Stick on his back a mustard plaster.” A natural rendition of the line admits of the following phonological grid representation:

18 CURRENT MUSICOLOGY

MWFR 3 requires four L(0) beats, which, according to MWFR 4, must be separated by one L(-1) beat. The only organization fulfilling this condition is:

Keep him there and make him bail her.
 * * * * * L(-1)
 * * * * * L(0)

Assigning appropriate durational values correctly completes the derivation:

Keep him there and make him bail her.

 * * * * * L(-1)
 * * * * * L(0)

Next we derive a setting for verse eight, a line having only two accented syllables:

x line (3)
 x line (2)
 x x line (1)
 x x x x x x x line (0)
 Keel - haul him till he's so - ber.

Proceeding from left to right, we assign next L(0) beats as follows:

Keel - ____ ____ ____ ____ so - ____.
 * * * * * L(0)

Next we distribute L(-1) beats from left to right to the remaining syllables:

Keel - haul him till he's so-ber.
 * * * * * L(-1)
 * * * * * L(0)

Now we assign the remaining L(0) beats required by MWFR 3. No possible assignment of L(0) beats can be derived that respects the binary spacing condition MWFR 4, as can be seen in (10).

(10a) Keel - haul him till he's so-ber.
 * * * * * L(-1)
 * * * * * L(0)

(10b) Keel - haul him till he's so-ber.
 * * * * * L(-1)
 * * * * * L(0)

An L(-1) beat must be interposed to achieve metrical well-formedness. We need to propose a rule that allows for the interposition of an additional L(-1) beat in particular environments. As has been noted by phonologists for some time,¹¹ two adjacent, or nearly adjacent, stressed syllables present a “stress clash,” which tends to be resolved by the speaker’s application of the “rhythm rule”: primary stress is deleted from the first of the two adjacent stressed syllables and shifted leftwards to a previous stressed syllable. Well-known examples of this phenomenon are shown in (11).

(11) The Rhythm Rule

		x			x	line (3)
	x	x		x	x	line (2)
x	x	x		x	x	line (1)
x	x	x		x	x	line (0)
thirteen	+ men	<i>becomes</i>	thirteen	men		
		x			x	line (3)
	x	x		x	x	line (2)
x	x	x		x	x	line (1)
x	x	x	x	x	x	line (0)
Mississippi	+ Mike	<i>becomes</i>	Mississippi	Mike		

In textsetting, stress clashes may be reduced in certain contexts by application of the rhythm rule. That is, the stress grid is altered and a setting is derived based on the altered stress grid. More common, however, is the alleviation of the clash by actual temporal separation of the two adjacent stressed units. In (10b), the stress clash between “keel” and “haul” provides the opportunity of achieving a well-formed grid by the insertion of an additional L(-1) beat between the two syllables. We therefore propose the following:

Beat Addition Rule: An L(-1) beat may be interposed between two stressed syllables in order to achieve a well-formed grid.

The parenthesized asterisk at level L(-1) is interposed in the environment specified in the Beat Addition Rule, achieving the desired setting:

Keel	- haul	him	till	he's	so	- ber.	
()	*	*	*	*	*	*	L(-1)
*	*		*		*		L(0)

¹¹ Morris Halle and Noam Chomsky, *The Sound Pattern of English* (New York: Harper and Row, 1968), 117.

“metrically rigid”¹³—whether specifically poetic as in the case of nursery rhymes and jump-roping songs, or musical as in the case of rap music—modified versions of the algorithm can be advanced that generate more or less plausible normative settings. Also worth investigation is the role of these normative settings in constraining the choices of composers, either positively or negatively, in confirming or overturning listeners’ expectations for the “most natural” correspondence of text and tune.

Appropriately extended versions of the algorithm can assign metrical structure to texts that are not normally intoned in a metrically rigid fashion. This inquiry may therefore shed light on prosodic idioms that have remained problematic from the standpoint of traditional prosodic theory. Foremost among these are blank verse as practiced by Donne, Milton and Shakespeare, as well as Hopkins’ highly abstruse Sprung Rhythm, discussed recently by Paul Kiparsky.¹⁴ A further expansion might involve an application to the intonational structure of phrases within normal speech, in order to explain a speaker’s highly subtle intuitions with respect to the “rhythms of speech” that form a significant component of his unconscious knowledge of his language.

ABSTRACT

Formalisms borrowed from generative music theory and generative phonology are employed to represent abstract structures underlying textsetting. An algorithm is then advanced which is shown to produce appropriate settings for a well-known strophic song.

¹³ R.T. Oerhle, “Temporal Structures in Verse Design,” in *Phonetics and Phonology, Rhythm and Meter*, ed. Paul Kiparsky and Gilbert Youmans (San Diego: Academic Press, 1989), 87–119.

¹⁴ Paul Kiparsky, “Sprung Rhythm,” in *Phonetics and Phonology*, 305–40.