

# The Sounds of Poetry Viewed as Music

FRED LERDAHL

*Department of Music, Columbia University, New York, New York 10027, USA*

**ABSTRACT:** An extended parallel is developed between musical and prosodic structures, using the author's cognitively oriented music theory and recent work in generative phonology. For illustration, the sounds of a short poem by Robert Frost are treated entirely in musical terms. The poem is assigned a phonological stress grid and then musical grouping and meter. These structures enable a durational realization. Phonological stress also helps assign the poem's normative melodic contour. Finally, the similarities and differences in sound repetition are given hierarchical structure by means of musical prolongational theory. These formal parallels suggest a corresponding realization in brain localization and function. Evidence from the neuropsychological literature is cited in support of this view. The picture emerges that grouping, meter, duration, contour, and timbral similarity are mind/brain systems shared by music and language, whereas linguistic syntax and semantics and musical pitch relations are systems not shared by the two domains.

**KEYWORDS:** Poetry and music; Music and poetry; Phonology, generative; Frost, Robert

## INTRODUCTION

Comparisons between music and language have traditionally been couched in terms of syntax or rhetoric.<sup>1</sup> The more substantive parallels, however, are those between musical structures, on the one hand, and phonological and prosodic structures, on the other, and derive from the fact that both music and language consist of sounds organized in time. These parallels are explored by the development of a representational and partly computational account of the sounds of a poem treated entirely as music. This approach builds on work from previous collaborations and consultations.<sup>2–5</sup> As I hope to show, the analysis is suggestive for the neurobiology of music.

## A MUSICAL-POETIC ANALYSIS

Consider in some detail the first two lines of Robert Frost's short lyric, "Nothing Gold Can Stay," given in its entirety in FIGURE 1. Grouping structure, a fundamental component of music theory, segments a musical surface hierarchically into motives, phrases, and sections. Phonologists have developed the comparable concept of the prosodic hierarchy.<sup>6</sup> Hence, linguistics now offers two kinds of tree structure, a syntactic and a prosodic one. Even when the two coincide in their segmentations, their

Address for correspondence: Dr. Fred Lerdahl, Department of Music, Columbia University, New York, NY 10027. Voice: 212-854-1295.  
fred@music.columbia.edu

Nature's first green is gold,  
 Her hardest hue to hold.  
 Her early leaf's a flower;  
 But only so an hour.  
 Then leaf subsides to leaf.  
 So Eden sank to grief.  
 So dawn goes down to day.  
 Nothing gold can stay.

FIGURE 1. "Nothing Gold Can Stay," by Robert Frost.

meanings differ. Syntactic trees represent parts of speech and syntactic phrases; phonological trees represent groupings of spoken sound.

At the lowest levels of the prosodic hierarchy are the syllable and the word. Words are categorized as either content words, which carry major semantic content, or function words, which mainly fill a syntactic role. For example, in the first line of the poem, "Nature's first green is gold," all the words are content words except "is." At the level above the word is the clitic group, which either coincides with a content word or combines content and function words that, together, sound as if they were one word. In the latter case, the stressed syllable of the content word is the clitic host, the sound to which the others are drawn. In "Nature's first green is gold," "is gold" sounds as one word, with "gold" as the clitic host. Likewise in the second line, "Her hardest hue to hold," the content words are "hardest," "hue," and "hold"; "Her hardest" and "to hold" are clitic groups, with "hard-" and "hold" as clitic hosts. Above the clitic group is a grouping of words, called the phonological phrase, for example, "first green." Above that is the intonational phrase, which conveys the melody of speech. The entire line, "Nature's first green is gold," is an intonational phrase. Finally, there is the utterance, which usually corresponds to a sentence. "Nature's first green is gold, / Her hardest hue to hold" is an utterance.

FIGURE 2 displays the grouping analysis of the first two lines of the poem. The syllable and word levels are omitted for convenience. The bracketing categories are indicated in boldface: **C** = clitic group, **P** = phonological phrase, **I** = intonational phrase, and **U** = utterance. A category can repeat from one level to the next. This representation abstractly resembles a parsing of a musical phrase into subgroups and

[ <b>C</b> Nature's]	[ <b>C</b> first]	[ <b>C</b> green]	[ <b>C</b> is gold]	[ <b>C</b> Her hardest]	[ <b>C</b> hue]	[ <b>C</b> to hold]
[ <b>P</b> Nature's]	[ <b>P</b> first	green]	[ <b>P</b> is gold]	[ <b>P</b> Her hardest	hue]	[ <b>P</b> to hold]
[ <b>I</b> Nature's	first	green	is gold]	[ <b>I</b> Her hardest	hue	to hold]
[ <b>U</b> Nature's	first	green	is gold	Her hardest	hue	to hold]

FIGURE 2. Bracketed, leveled representation of the prosodic analysis of the first couplet.

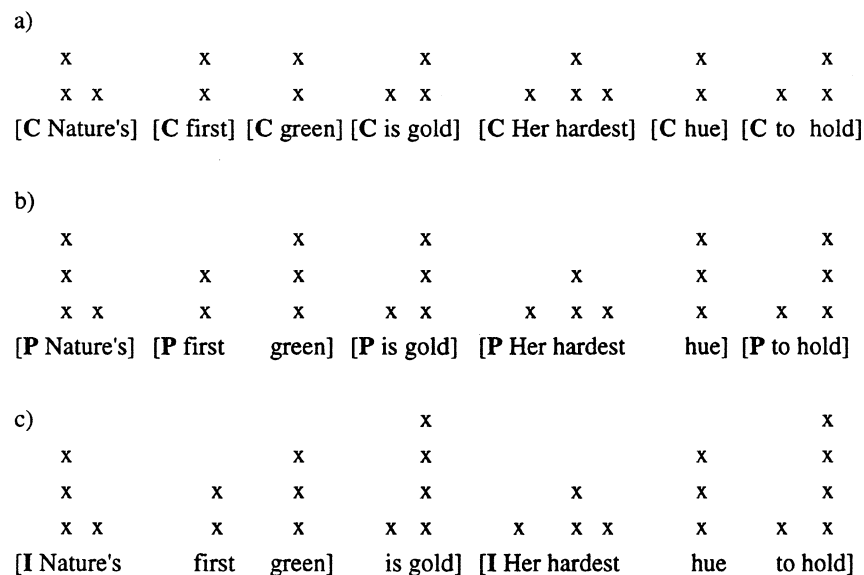
motives. As in music, pauses are more likely to occur between, rather than within, these segments, with greater pauses between superordinate boundaries.

Now consider linguistic stress, whose musical equivalent is phenomenal accent. In both, the perception, whether of a syllable or a pitch event, is one of relative sonic prominence within its immediate context. Linguists have proposed a so-called metrical grid to represent linguistic stress in words and phrases.<sup>7</sup> From a musical perspective, this is a misnomer; it should be called a stress grid, for its purpose is to represent not hierarchical periodicities but hierarchical patterns of stress. Within a word, such a pattern is given lexically, as in the first word of the poem, “Náture’s” (one would not say “Natúre’s”). Within a phrase, the stress pattern obeys the nuclear stress rule, which, after the categorization of words into content and function words, assigns global stress to the strongest syllable of the last content word of a prosodic unit.<sup>8</sup> Thus, we say “first gréen,” rather than “first green,” because “green” is the last stressed syllable of its phonological phrase. This principle can be overridden by the rhythm rule, which ameliorates clashes in stress between adjacent and nearby syllables. According to nuclear stress, one says “hardest húa,” but we might prefer to adjust it to “hárddest hue,” so that the nearby words “hue” and “hold” are not comparably stressed. The rhythm rule implicitly invokes metrical periodicity: strong beats should be more or less evenly spaced. At this level, then, the choice is between “Her hardest húa to hólđ,” which obeys nuclear stress but challenges periodicity, and “Her hárddest hue to hólđ,” which violates nuclear stress in order to distribute the stresses. Finally, nuclear stress can be overridden by nonnormative focus on a particular word, to bring out a semantic nuance. Hence, if we say “Nature’s first green is gold,” the implication is that nature’s second green is not gold. Generally, however, nuclear stress holds.

Setting the rhythm rule and focus to one side, a stress grid is constructed by cyclic generation of stresses from level to level in the prosodic hierarchy, observing nuclear stress starting at the phonological grouping level. As illustrated in FIGURE 3a, the procedure is first to assign an *x* to every syllable. Second, assign an *x* in the next row to a lexically relatively stressed syllable in a polysyllabic word or to the host of a clitic group. For the third row, shown in FIGURE 3b, add an *x* for the nuclear stress at the level of the phonological phrase. Finally, in FIGURE 3c, add an *x* for the nuclear stress at the level of the intonational phrase. There is little point in continuing to yet another level, for, as in music, distinctions in linguistic stress attenuate over long spans.

FIGURE 4 recasts the stress pattern in FIGURE 3 into an equivalent music-reductional format. Level *d* brackets the syllables and corresponds in content to the bottom row of the stress grid in FIGURE 3. Level *c* reduces out the syllables and monosyllabic words that in FIGURE 3 are assigned only one *x*. Similarly, levels *b* and *a* keep only those units that carry three and four *x*s, respectively. This layout is the same as that in musical time-span reduction, which selects relatively important pitch events at successively larger levels of rhythmic segmentation. These relationships could alternatively be displayed by a tree representation with right and left branching, again as in corresponding music theory.

Now turn to meter. In both music and poetry, metrical structure consists of hierarchically related periodicities inferred from the signal. It might be objected that periodicities in language, unlike those in the musical case, do not really exist because



**FIGURE 3.** Assignment of the stress grid in conjunction with the prosodic hierarchy applied to the couplet.

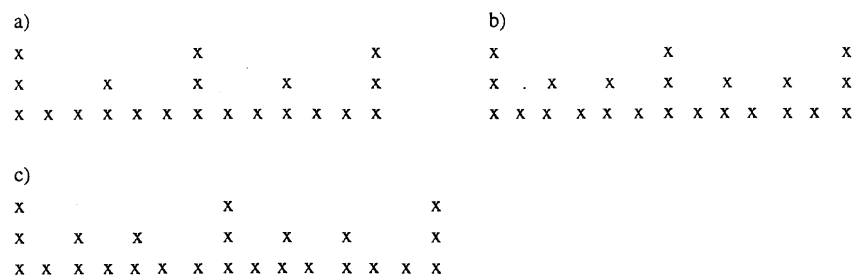
level *d*: [Na][ture's] [first] [green] [is][ gold] [Her hard][est] [hue] [to] [hold]  
 level *c*: [Na- ] [first] [green] [ gold] [ hard- ] [hue] [ hold]  
 level *b*: [Na- ] [ green] [ gold] [ hue] [ hold]  
 level *a*: [ gold] [ hold]

**FIGURE 4.** The stress pattern in FIGURE 3 put in reductional format.

syllabic and phrasal durations are so variable. Yet, it would be misleading to say that musical durations are invariable. Expressive musical performance depends on deviations from isochrony. Like meter itself, temporal precision is a mental construct. While it is true that durations in verse are usually more variable than those in music, many poetic idioms, from nursery rhymes to sophisticated traditions, demonstrate considerable regularity.<sup>9</sup> As in music, these verbal idioms approach periodicity as a framework against which expressive deviations take place.

Once periodicity is understood as a relative matter, poetic and musical meter may appropriately be regarded as formally and cognitively equivalent. This has, in essence, long been recognized by music theorists in their use of prosodic foot notation for representing musical meter. Now, however, it is more usual in music theory to employ a grid notation, which represents the metrical hierarchy directly and does not confuse features of meter with those of grouping. A similar consensus has emerged in phonological theory.



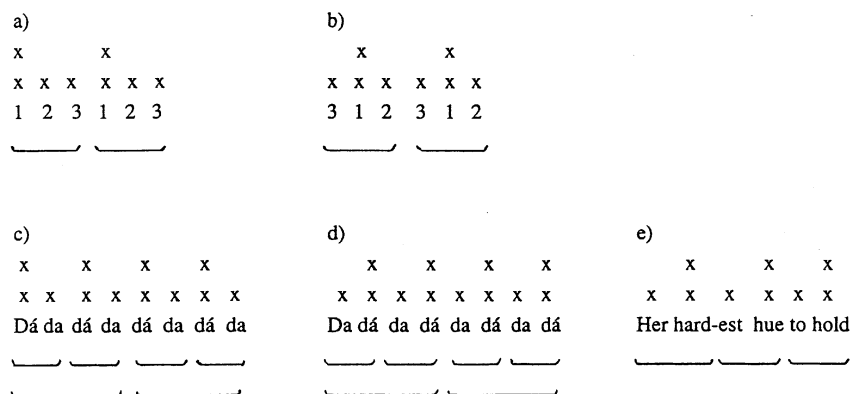


**FIGURE 5.** Three metrical grids: (a) 6/8 meter; (b) 3/4 meter; (c) 7/8 meter (2 + 2 + 3).

A poetic or musical meter exists when the perceiver infers conceptually regular levels of beats from the signal, so that a beat that is strong at one level is also a beat at the next larger level. The perceptually most prominent metrical level is called the *tactus*; it occurs at a moderate tempo, is usually at an intermediate level within a grid, and is often more restricted in regularity than are smaller or larger levels. Beats at a given level are two or three beats apart at the next smaller level. (Even lines in pentameter have intermediate strong beats, resulting in patterns of two and three.) Some poetic or musical idioms, while allowing a mixture of two or three beats apart across levels, discourage the mixture of two or three strong beats apart within a level. Thus, European classical tonal music uses compounds of two and three across levels, as in FIGURE 5a and b, but it regards as atypical the grid in FIGURE 5c, which alternates two and three beats within a single level. Other musical cultures, as those of the Balkan countries, would take FIGURE 5c as normative. The important point is not that different cultures prefer different metrical grids, but that the possible combinations of two and three, across or within levels, are very small. Cultural variation on poetic and musical meters is intrinsically limited.

In constructing a metrical grid, the perceiver seeks an optimal fit between the stresses provided by the acoustic signal and the culturally available repertory of metrical grids. To hear a musical phrase in 3/4 time or a poetic line in pentameter is to select that particular grid because, within the repertory, that grid gives the minimal number of conflicts between stress and beat. When the fit is imperfect, the effect is one of syncopation or, to use terms from traditional scansion, of inversion, substitution, or truncation. If the fit is seriously imperfect, the perceiver switches to another available grid or perhaps abandons metrical understanding altogether.

The interaction of grouping and meter causes intuitions of upbeat and afterbeat in relation to a downbeat. The meter is 3/4 in FIGURE 6a and b, but the two differ in that FIGURE 6a has afterbeats, while FIGURE 6b has an upbeat and an afterbeat. The terminology of traditional scansion recognizes this upbeat–afterbeat distinction. In current notation, FIGURE 6c and d represents the contrasting groupings of trochaic and iambic tetrameter, respectively. Traditional foot scansion becomes troublesome, however, in cases where the presumed foot contradicts the prosodic hierarchy. Supposing our lyric to be in iambic trimeter, in the line, “Her hardest hue to hold,” the foot-grouping of “-est” is not with its own word stem, “hard-,” but counterintuitively with “hue,” as shown in FIGURE 6e. It is unclear in such cases what function the foot



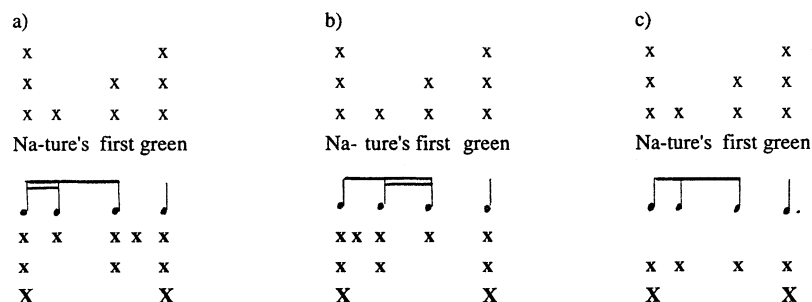
**FIGURE 6.** Interaction of grouping and meter: (a) 3/4 meter with afterbeats; (b) 3/4 meter with an upbeat and an afterbeat; (c) trochaic tetrameter; (d) iambic tetrameter; (e) conflict between word and foot boundaries at "hardest."

serves. In the present view, iambs, trochees, anapests, and so on arise not from a pre-determined schema but directly from the prosodic groupings of beats. Moreover, the concept of the foot lacks a musical counterpart. While recognizing its historical practice, let us ignore the foot as a unit of rhythmic analysis.

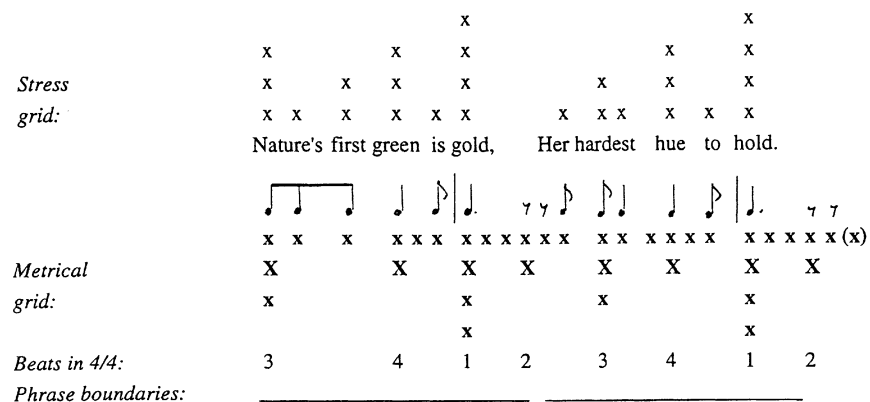
Having set to one side the iambic dimension of our poem, I would now suggest that it is not in trimeter either. It is in truncated tetrameter, because of the punctuated pauses at the end of each line and, more elusively, by virtue of the historical style that it evokes. Many short lyrics, such as those of Emily Dickinson, employ so-called common meter, comprising couplets in a four-plus-three beat pattern with an implied silent final beat: da-dá, da-dá, da-dá, da-dá; da-dá, da-dá, da-dá, ---. Common meter is a variant of standard tetrameter, the final silent beat reinforcing the grouping structure. Our poem is, in turn, a variant of common meter. Frost uses it here because of the ineluctable decline that is the poem's theme. Every line falls into silence.

Given this paradigmatic meter, the first two lines of the poem receive the structural description in FIGURE 7. Above the lines appears the stress grid, taken from Figure 3c. Beneath, in boldface and inverted for visual convenience, is the metrical grid for the tactus and larger levels, showing four basic beats per line and their relative metrical strengths. The tactus is represented by uppercase Xs. To clarify the grid notation, the beats are numbered as two 4/4 measures, grouped into two intonational phrases. Only stresses having three or four xs count for establishing the linchpin of the tactus. In accordance with the greatest stresses on "gold" and "hold," these syllables are assigned the strongest beats. The periodicity of the metrical grid requires that "Na-" and "hard-" receive the next strongest metrical position, with "green," "hue," and the silent beats getting the weaker beats. This treatment creates a slight syncopation at "hue." Observe that the strongest metrical beats take place at the end of each line, creating an out-of-phase relationship between the grouping boundaries and the spans between the strongest beats. This is a consequence of nuclear stress and is characteristic of poetry in English. In classical tonal music, by contrast, the strongest beat usually comes at or near the beginning of a phrase. For its ending





**FIGURE 8.** Duple and triple subtactus metrical realizations for the beginning of the first line.



**FIGURE 9.** Addition of the subtactus level to FIGURE 7, with translation into musical durations.

If this derivational procedure is carried out for the entire poem, the rhythmic derivation is as in FIGURE 10, cast in a 12/8 meter with two-beat anacrusis for each line. Minor adjustments with duple divisions appear in the second, third, and fourth lines; these have to do with syllable length, a topic I pass over here (compare "hard-est" in FIGURE 9).

We turn now by way of stress to contour. Three acoustic variables contribute to stress: intensity (or loudness), pitch height, and duration.<sup>11</sup> While duration is an independent variable, intensity and pitch height covary, for the muscles that enable greater intensity also tend to raise pitch. Even when pitch height is artificially isolated, relative height induces the perception of relative stress. Thus, pitch height is a strong projector of stress. Since pitch height is a matter of contour, stress patterns can provide a basis for the derivation of the melody of speech. There are, however, many acceptable contours for an intonational phrase or utterance, each having its own nuance. If our concern were grammaticality, in keeping with much generative



FIGURE 10. Durational reading of the poem.

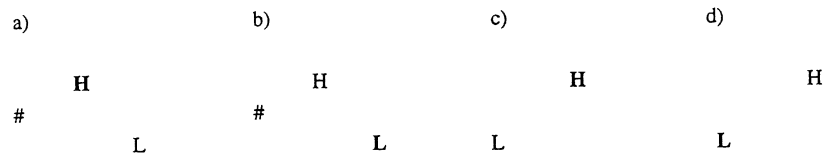


FIGURE 11. Four normative contour frameworks.

linguistics, well-formedness constraints would underdetermine the options for contour realizations. A different goal, and one congenial with a musical perspective, is to ask whether a grammatical contour is normative or not. We need then derive only normative contours from stress grids, letting the other grammatical contours be seen as nuanced variants.

To pursue this question, let us posit the restriction of contour features to a binary classification of high and low, designated by “H” and “L,” and apply it to the two most stressed syllables in an intonational phrase.<sup>12</sup> The more stressed of the two is shown in boldface. This is a reductional system in a musical sense: dominating **H** or **L**, then secondary H or L, then all the rest of the syllables in the phrase. Suppose, further, that H does not occur at the left edge of the phrase, that is, that a normative

intonational phrase does not begin with its most important peak. There is no symmetrical constraint for H at the right edge or for L at either edge. The four logical possibilities are given in FIGURE 11, with “#” standing for syllables to the left of the first H. The prototypical declarative contour is shown in FIGURE 11a and b: a rise and fall whose musical analogue is the tensing-relaxing pattern of a standard tonal phrase. The difference between the two is whether H or L dominates. In FIGURES 11c and d, H is to the right of L; this is the typical shape of an interrogative.

The claim that a higher pitch causes stress would seem to disqualify FIGURE 11b and d as normative, because in these cases **L** dominates. As a consequence of nuclear stress, however, **L** can indeed function in the position of the last clitic host of an intonational phrase. In other words, **L**'s location at the close of a major phrase boundary compensates for its lower pitch. Hence, FIGURE 11b is normative. By the same token, FIGURE 11d is not normative, because in this case **L** is not in a location to benefit from major nuclear stress. As there are no interrogatives in our poem, we can also ignore FIGURE 11c. This leaves FIGURE 11a and b as the remaining options. This very restricted repertory will provide the framework for mapping stress importance onto contour peaks and valleys.

The stress analysis in FIGURE 3 assigns three *xs* for the most stressed syllable in each phonological phrase and four *xs* for the most stressed syllable in each intonational phrase, the latter occurring at the end of the line. Hence, as shown in FIGURE 12a, the intonational contour falls in the pattern of FIGURE 11b. The two lines together, however, form an utterance, leading to the assignment in FIGURE 12b of H and L to four *xs* at this level, over “gold” and “hold,” respectively. The assignment for “gold” is thereby reversed from L in FIGURE 12a to H in FIGURE 12b. This step does not yet determine whether H or L in FIGURE 12b is dominating; for reasons beyond the present purview, **H** is chosen. The next step is to combine the shapes at the utterance and intonational levels. In FIGURE 12a, “green” received an H, an assignment that is no longer preferred because it would contradict the **H** in FIGURE 12b. As shown in FIGURE 12c, the H on “green” becomes an L within its intonational phrase; the brackets signify its subordinate status. The H on “hue” remains an H, for it combines at the intonational level with an **L**, although it too is placed in brackets to indicate its lesser hierarchical position. This is the scaffolding within which more local contour shapes take place.

In moving from the abstract H-and-L representation to realized contours, it is useful to posit discrete levels of relative pitch height. Four levels appear to be optimal: three do not afford enough distinctions, and more than four would be fussy in view of the limited ability to categorize degrees of stress. The H-and-L representation is mapped onto these four levels. On the basis of the salience of registral extremes, **H** is placed in the highest slot and L is placed in the lowest slot. FIGURE 13a performs this step for the first couplet, using two lines and two spaces for the four pitch levels and representing degrees of hierarchical importance by relative durational values (that is, these durational values signify not durations but structural importance, as is often done in musical reduction theory). FIGURE 13b appends pitch heights for each [H] and [L], so that each [H] or [L] is placed one slot lower than its adjacent **H**. These locations preserve the rising–falling shape of the contour schema in FIGURE 11a and b. The H-and-L notation can be dispensed with at smaller levels. Except where **L** or L occurs at the right edge of an intonational phrase, the contour reflects the shape of

a)

H						H					
						L					
						x					
x			x		x				x		x
x		x	x		x			x	x		x
x	x	x	x	x	x		x	x	x	x	x

Nature's first green is gold,      Her hardest hue to hold.

b)

H						L					
						x					
x			x		x				x		x
x		x	x		x			x	x		x
x	x	x	x	x	x		x	x	x	x	x

Nature's first green is gold,      Her hardest hue to hold.

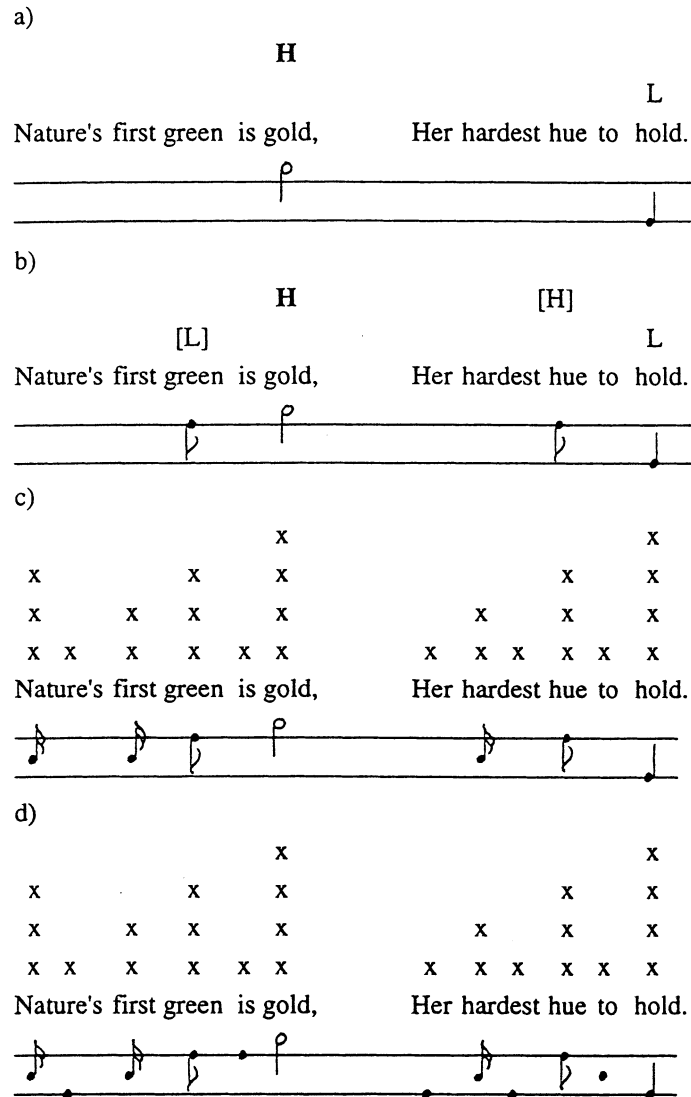
c)

H						[H]					
[L]						L					
						x					
x			x		x				x		x
x		x	x		x			x	x		x
x	x	x	x	x	x		x	x	x	x	x

Nature's first green is gold,      Her hardest hue to hold.

**FIGURE 12.** Global contour assignments for the first couplet.

the stress grid, with its peaks and valleys of *xs*. Thus, in FIGURE 13c the syllables with three and two *xs*, notated by sixteenth notes, appear one slot below the positions of the syllables that were already given three *xs* in FIGURE 13b. Then, in FIGURE 13d come the syllables with one *x*, represented by unstemmed noteheads. These syllables are cliticized and are typically spoken as schwas (the default neutral vowel for an unstressed syllable). They are placed one step below their clitic hosts, except when the latter are in the lowest slot, in which case they are located one step above the clitic



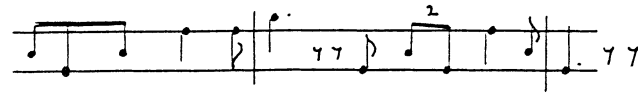
**FIGURE 13.** Assignment of pitch levels to the first couplet: (a,b) from the H-and-L analysis in FIGURE 12; (c) for unassigned syllables with three and two xs; (d) for syllables with one x.

host. Hence, “-ure’s” is one step below “Na-” in “Nature’s,” and “is” is one step below “gold,” but “to” is one step above “hold,” because the latter is in the bottom slot.

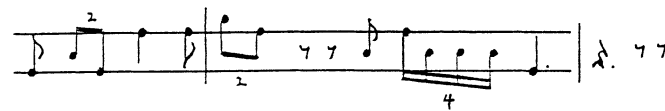
This derivation of contour is musical not only in its attention to normative shapes but also in its hierarchical treatment. Rather than generate contours from left to right, as if without projection or memory, global highs and lows are established according



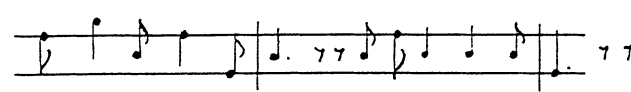
Nature's first green is gold, Her hardest hue to hold.



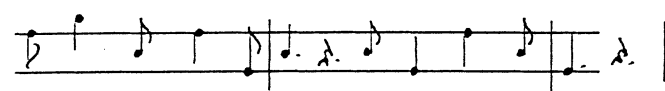
Her early leaf's a flower; But only so an hour.



Then leaf subsides to leaf. So Eden sank to grief.



So dawn goes down to day. Nothing gold can stay.



**FIGURE 14.** Meter and durations (from FIGURE 10) combined with a contour analysis of the entire poem.

to a few paradigmatic shapes, and the remaining syllables in the speech melody are filled in at successive levels of lesser stress. FIGURE 14 shows the result, combining the contour analysis of FIGURE 13 with the metrical/durational analysis of FIGURE 10, but now with these methods applied to the entire poem. Contained within this seemingly transparent musical notation are the structures of phonological stress, the prosodic hierarchy, the metrical grid, duration, and pitch height. The success of the derivation can be judged informally by its naturalness as a reading.

A usual prosodic analysis would have stopped by this point. From a musical perspective, however, there is yet another step to take, an analysis of the hierarchical patterns of recurrence of sounds. Traditional poetic analysis treats verbal recurrences such as rhyme, alliteration, and assonance as primitive sequential patterns: *aabb*, *abab*, and so forth. Music theory, by contrast, has a highly developed approach to recurrence in the form of prolongational structure. FIGURE 15 gives three kinds of prolongational connection: (1) strong prolongation or repetition, represented by a dashed slur; (2) weak prolongation or partial repetition, represented by a dotted slur;

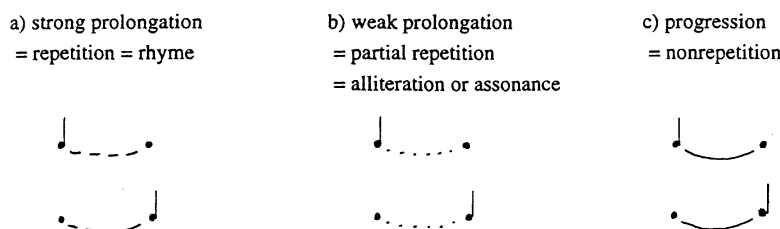


FIGURE 15. Repertory of prolongational connections.

Nature's first green is gold, Her hardest hue to hold.

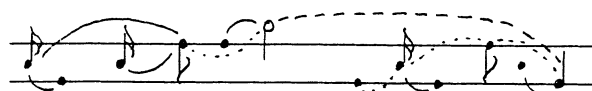


FIGURE 16. Prolongational structure for the first couplet. The contours are retained from FIGURE 14.

and (3) progression or nonrepetition, represented by a solid slur. In each case the superordinate element can precede or succeed the elaborating element.

Musical prolongational connections are derived through a combination of time-span segmentation and tonal pitch stability. Poetry incorporates the former through the prosodic hierarchy but lacks any counterpart to the tonal system. The better musical analogy, then, is to atonal music, for which event salience largely replaces event stability.<sup>13</sup> That is, when hearing atonal surfaces, and given the absence of acoustic or idiom-specific criteria to select stable from unstable events within hierarchical time spans, the listener tends to organize embellishing events in relation to events that have greater surface prominence within the framework of the time-span hierarchy. In phonological terms, relative surface prominence is equivalent to relative stress. Hence, prosodic prolongation can be viewed as a kind of atonal prolongation, in which the elements being prolonged are not pitch structures but degrees of timbral similarity mediated by stress within the prosodic hierarchy. (This approach extends beyond that of Lerdahl,<sup>14</sup> in which synthesized vowels are related prolongationally by similarity but without incorporating nested patterns of stress.) The three prolongational categories continue intact: the repetition of rhyme equals strong prolongation; the weaker repetition of alliteration or assonance equals weak prolongation; and syllabic nonrepetition equals progression.

The syllables connected in the prolongational analysis of the first couplet follow the parsing of the prosodic hierarchy in FIGURE 3 and the hierarchical contour in FIGURE 13d. Each connection forms a progression, weak prolongation, or strong prolongation, according to the degree of repetition of sound. FIGURE 16 illustrates, retaining the structural notation of FIGURE 13d. In the first line, “-ture’s” attaches as a progression from “Na-,” “first” to “green,” and “is” to “gold.” At the next level, “Na-” attaches to “green,” because both have three *xs*. Next come the attachments of

[H]s and [L]s to Hs and Ls, and then of the latter to **Hs** and **Ls**. Thus, “green” attaches “gold” as a weak prolongation because of the alliteration between “green” and “gold.” The second line creates a similar pattern, but with more internal alliteration. Next, at the level of the couplet, the two syllables with four *xs*, “gold” and “hold,” form a rhyming strong prolongation. Note it is timbral similarity rather than pitch that is being connected prolongationally, for “gold” is in the highest pitch category while “hold” is in the lowest. The resulting graph shows not isolated instances of alliteration and rhyme, as in traditional prosodic analysis, but the richer relationship of partial repetitions nested within superordinate rhymes—“green” is to “gold” as “hue” is to “hold.” Progressions between dissimilar syllables occur as relatively local connections. This approach can be extended to the poem as a whole.

This account of prosodic prolongational structure has skipped some derivational stages involving technicalities of the theory of musical reduction. The main point is that we can indeed transplant most of the mechanisms of the musical reduction theory to the poetic realm.

### CONNECTIONS

Let us step back from the derivation of the Frost analysis and assume that the music-analytic methods outlined here have validity for modeling the sonic organization of poetry in general. To what extent do these methods extend to the analysis of ordinary speech? Briefly, ordinary speech possesses phonological stress, the prosodic hierarchy, and contour; so the application of these components should continue intact. On the other hand, the phonological stresses of ordinary speech are irregular enough to inhibit the inference of a metrical grid. Similarly, ordinary speech lacks significant patterns of syllabic repetition, so that the sense of its timbral prolongational structure is weakened. These attenuated perceptions are reflected in the model by weakened metrical and prolongational derivations.

A complementary approach to studying the relationship between music and poetry would be to develop a rigorous theory of textsetting.<sup>15</sup> That listeners have strong intuitions about whether a poetic line is well- or ill-set to music is in itself an indication of the parallels between the two media. Generally, the stronger the correlation between the common formal structures of poetry and music, the more natural the textsetting. That is, a natural textsetting tends to match up the stresses, metrical positions, durations, phrasings, contours, and formal patterns of a poem and its musical realization. The idiom-specific traditions of *Sprechstimme* in early twentieth-century German musical expressionism and of hip-hop in current American popular music show how far music and language can merge along these dimensions.

The overall thrust of the Frost analysis is that the mental representation of the sounds of metrical, rhymed poetry and of music, whether texted or not, shares a good deal more organization than has usually been supposed. Perhaps this commonality has not been sufficiently recognized because in certain basic respects the two media are unlike: poetry has words and phrases with propositional meaning; music has a hierarchical, multidimensional pitch space.<sup>16–18</sup> The analysis of the Frost reflects these differences by not referring to the poem’s imagery or meaning and by not invoking musical concepts such as specific pitches and intervals, scales, harmony, key, voice leading, counterpoint, or tonal tension and relaxation.

<i>exclusively musical structures</i>	<i>common structures</i>	<i>exclusively linguistic structures</i>
pitches & intervals	durational patterns	syntactic categories & relations
scales	grouping (prosodic hierarchy)	word meaning (lexicon)
harmony	stress (contextual salience)	semantic structures (sense, reference, entailment)
counterpoint	metrical grids	phonological distinctive features (etc.)
tonality	contour	
pitch prolongations	timbral prolongations	
tonal tension & attraction		

**FIGURE 17.** Hypothesized brain organization of musical and linguistic structures.

It seems reasonable to suppose that well-founded theoretical parallels and distinctions have counterparts in brain structure and function. Appropriate theory can thus guide and interpret neurological findings. We may inquire, then, to what extent the commonalities and differences discussed here are instantiated in brain structure and function. The present analysis predicts, as one would expect, that there is not a single music area and a single language area, but that musical and linguistic functions are spread out in different structural and functional components. It further predicts that those brain modules that process rhythm, contour, and timbral relationships are the same in music and language, while those that process purely pitch structures and purely linguistic structures occupy different parts of the brain. FIGURE 17 illustrates.

Current neuropsychological evidence offers some support for this view. Studies of patients with brain lesions that affect musical and linguistic processing indicate that musical and phonological contours are processed in the same area of the right hemisphere but that musical pitch structures are processed in the left hemisphere, with a one-way channel from contour processing to intervallic pitch processing.<sup>19–21</sup> If so, individuals with deficits that inhibit their ability to process musical or linguistic contour are also unable to process specific pitch relations. This makes intuitive sense, for if a listener cannot tell up from down, how could he or she distinguish between a minor third and a perfect fourth? On the other hand, individuals unable to process musical pitch might still be able to process contour. Normal people who are musically tone-deaf do not speak in a monotone. (In this connection, tone languages such as Chinese depend not on fixed pitch categories and intervals but on relative pitch height and contour.<sup>22</sup> Thus, a person could remain competent in speaking a tone language despite an inability to distinguish musical pitch relationships.)

Neuropsychological evidence also suggests a bifurcation in musical processing between pitch and rhythm. There is partial justification for this from music theory, in which grouping and meter are distinct components, from both tonal reduction and tonal pitch space.<sup>2,17</sup> Support for this division comes not only from neuropsychology but also from experiments in music cognition.<sup>23</sup> In the details, however, the story is bound to be more complex. From a music-theoretic perspective, pitch is one of the inputs to the derivation of grouping and meter. For example, a tonal cadence signals

a grouping boundary, and harmonic rhythm is a strong determinant of meter. Similarly, the derivation of pitch reduction depends on the position of pitches and chords within nested time spans based on grouping and meter. On the other hand, tonal space itself—the system (or systems) of pitches, intervals, scales, chords, tonal regions, tonal attractions, degrees of sensory and cognitive consonance, and degrees of mutual proximity and stability—remains, at a theoretical level, truly independent of duration, grouping, and meter. FIGURE 17 projects this overall picture: except for contour and timbral prolongations, all of the items listed under “common structures” belong to the domain of rhythm, which music and poetry share, and most of the items listed under “exclusively musical structures” belong to tonal space.

Note that the category “exclusively linguistic structures” in FIGURE 17 includes not just syntactic and semantic features but some phonological aspects as well, such as distinctive features.<sup>8</sup> Moreover, the perceptual organization of speech may not fully submit to a general, Gestalt-based auditory account.<sup>24</sup> These qualifications nevertheless do not undermine the convergence between theoretical and empirical treatments for the domains listed under “common structures.”

Presumably these exclusive and common structures are a consequence of human evolution. In this view, the roots of music and language are the same, in the form of premusical and prelinguistic communicative and expressive auditory gestures involving shapes of duration, stress, contour, timbre, and grouping. We still communicate with infants and higher mammals in this manner. I believe that these elementary shapes lie at the basis of expressive utterance in language and of musical expression (an idea that goes back at least to Rousseau<sup>25</sup>). With evolution came specialization: music and language diverged in their most characteristic features, pitch organization in music and word and sentence meaning in language. Poetry straddles this evolutionary divergence by projecting, through the addition to ordinary speech of metrical and timbral patterning, its common heritage with music.

### ACKNOWLEDGMENT

I am grateful to the phonologist Bruce Hayes for advice on aspects of this study.

### REFERENCES

1. WINN, J. 1981. *Unsuspected Eloquence: A History of the Relations between Poetry and Music*. Yale University Press. New Haven.
2. LERDAHL, F. & R. JACKENDOFF. 1983. *A Generative Theory of Tonal Music*. MIT Press. Cambridge, MA.
3. LERDAHL, F. & J. HALLE. 1991. Some lines of poetry viewed as music. *In Music, Language, Speech, and Brain*. J. Sundberg, L. Nord & R. Carlson, Eds. Wenner-Gren International Symposium Series. Macmillan. London.
4. LERDAHL, F. & X. CHABOT. 1991. A Theory of Poetry as Music and Its Exploration through a Computer Aid to Composition. *Proceedings of the International Computer Music Conference*, 1991.
5. HAYES, B. Personal communication.
6. HAYES, B. 1989. The prosodic hierarchy in poetry. *In Phonetics and Phonology: Rhythm and Meter*. P. Kiparsky & G. Youmans, Eds. Academic Press. New York.
7. LIBERMAN, M. & A. PRINCE. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8: 249–336.

8. Chomsky, N. & M. Halle. 1968. *The Sound Pattern of English*, 2nd printing. MIT Press. Cambridge, MA.
9. OEHRLE, R.T. 1989. Temporal structures in verse design. *In* *Phonetics and Phonology: Rhythm and Meter*. P. Kiparsky & G. Youmans, Eds. Academic Press. New York.
10. BECKER, J. & A. BECKER. 1979. A grammar of the musical genre *Srepegan*. *J. Music Theory* **23**: 1–44.
11. HANDEL, S. 1989. *Listening*. MIT Press. Cambridge, MA.
12. PIERREHUMBERT, J. 1980. *The Phonology and Phonetics of English Intonation*. MIT Ph.D. dissertation.
13. LERDAHL, F. 1989. Atonal prolongational structure. *Contemp. Music Rev.* **4**: 65–87.
14. LERDAHL, F. 1987. Timbral hierarchies. *Contemp. Music Rev.* **1**: 135–160.
15. HALLE, J. & F. LERDAHL. 1994. A generative textsetting model. *Curr. Musicol.* **55**.
16. KRUMHANS, C.L. 1990. *Cognitive Foundations of Musical Pitch*. Oxford University Press. New York.
17. LERDAHL, F. 1988. Tonal pitch space. *Music Percept.* **5**: 315–350.
18. LERDAHL, F. 1996. Calculating tonal tension. *Music Percept.* **13**: 319–363.
19. PERETZ, I. 1993. Auditory agnosia: a functional analysis. *In* *Thinking in Sound: The Cognitive Psychology of Human Audition*. S. McAdams & E. Bigand, Eds. Oxford University Press. Oxford.
20. PERETZ, I. 2001. Le cerveau musical. *In* *Enciclopedia Einaudi della musica*. J.J. Nattiez, Ed. In press.
21. PATEL, A.D. & I. PERETZ. 1997. Is music autonomous from language? A neuropsychological appraisal. *In* *Perception and Cognition of Music*. I. Deliège & J. Sloboda, Eds. Psychology Press. Hove, UK.
22. YUNG, B. 1991. The relationship of text and tune in Chinese opera. *In* *Music, Language, Speech, and Brain*. J. Sundberg, L. Nord & R. Carlson, Eds. Wenner-Gren International Symposium Series. Macmillan. London.
23. PALMER, C. & C.L. KRUMHANS. 1987. Independent temporal and pitch structures in determination of musical phrases. *J. Exp. Psychol. Human Percept. Perform.* **13**: 116–126.
24. REMEZ, R.E., P.E. RUBIN, S.M. BERNIS, *et al.* 1994. On the perceptual organization of speech. *Psychol. Rev.* **101**: 129–156.
25. ROUSSEAU, J.-J. 1760. *Essai sur l'origine des langues*. *In* *On the Origin of Language*. 1986. J. H. Moran & A. Gode, trans. University of Chicago Press. Chicago.