Constituency Matching in Metrical Texts

John Halle
Dept. of Music
Yale University
1/21/04

0. Introduction

Musical text setting and metrical poetry can be characterized as an attempt by composers and poets to match metrical positions and syllables in an aesthetically appealing manner. It is generally agreed that what lies behind these choices are more or less clear intuitions as to natural correspondences for position and syllabic quality. And it is also generally agreed that natural arrangements tend to be characterized by a correspondence of:

*Relatively strong positions and relatively stressed syllables and/or

*Relatively weak positions and relatively unstressed syllables.

Unnatural setttings and unmetrical lines tend to correspond

*Relatively strong positions and relatively unstressed syllables and/or

*Relatively weak positions and relatively stressed syllables.

These arrangements are precisely represented by the metrical grids below. Optimal settings tend to correspond S's (stressed syllables) with "beats" and unstressed syllables (U) with "off-beat" locations.

0.1a)

Hike		in	the	in-		come		tax.	
S		U	U	S		U		S	
X	X	X	X	X	X	X	X	X	X
X		X		X		X		X	
X				X				X	
				X					

Problematic or awkward settings result from "mismatches"-relatively stressed syllables assigned to weak locations (indicated in bold type) or relatively unstressed syllabes assigned to strong positions (italicized):

b)

Off		*the*	**free**-	*way*		ex-		*it*.*	
S		U	S	U		S		U	
X	X	X	X	X	X	X	X	X	X
X		X		X		X		X	
X				X				X	
				X					

There is, of course, much more to be said about what constitutes a mismatch. In particular, certain types of mismatches are more or less unnatural and some are routine in metrical verse, vocal music and rhythmic chants. The spondee, inverted foot, and other categories within traditional prosody were devised to account for the most common types. The ad hoc nature of these categories has given rise to numerous attempts to improve them from the 17th century onward (see Attridge (1982) and references therein for discussion). More recently, generative metrics has attempted to provide a formal explanation for the range of mismatches encountered within the corpus of metrical verse.

A significant lacuna not only of generative metrics but also much of traditional prosody consists in the assumption that the sole factor determining the intuitively based acceptability of lines are stress-meter correspondences of the sort identified in 0.1. While we recognize that these are surely significant in determining what constitutes natural text settings and acceptable lines of verse, we will argue that an additional factor, what we will call constituency matching, operates independently of (albeit to some extent interactively with) stress and meter We will provide evidence of the significance of constituency matching for text setting of intuitively derived constructs (such as those discussed in Halle (1997)). Intuitions relating to constituency will be seen to be highly relevant to composers' choices as reflected in the vocal literature. We also will suggest that these are likely to be significant for the closely allied set of intuitions governing metrical verse as well.

1. Low Level constituency matching

One reason for suspecting that an additional class of constraints is operative in text setting (and probably poetic meter) is based on the realization that it is possible to construct stress/meter correspondences which are highly stable from the standpoint of 0.1 a and b but remain relatively deviant. Included among these are the following examples presented in Halle's (1997) study of improvised text setting:

1.1

a)

X	X	X	X	X
X		X		X
X				X
Cat		in	the	hat.

b)

X	X	X	X	X
X		X		X
X				X
Buf-	fa-	lo		Bill

1.2

a)

X	X	X	X	X
X		X		X
X				X
Cat	in	the		hat.*

b)

X	X	X	X	X
X		X		X
X				X
Buf-		fa-	lo	Bill*

Clearly 1.1a and 1.1b are preferable to 1.2 a and b. Since both texts are composed of the same sequence of relatively stressed and unstressed syllables (SUUS), the sequence of stresses is equally well matched to metrical positions in each case. It is not obvious how the correspondence between textual vs. metrical accent can account for the perception of deviance.

If stress is not relevant to the determination of the acceptability of the metrical patterns assigned to the phrases above, it is reasonable to inquire as to what feature of language is being responded to by composers and others when they set texts to rhythms. The explanation suggested in Halle (1997) is that prosodic constituents (as described in Hayes (1989)) tend to be "grouped together" in musical text settings. The units defined by Hayes will be tend to be set off in vocal practice as separate "motifs," "motives," "phrases" or "groups."

What gives rise to the perception of separate musical constituency is the presence of a musical grouping boundary (Lerdahl and Jackendoff (1983)). Probably the most decisive cue for the assignment of a grouping boundary is a temporal gap between the attack points of two musical events. In relatively preferred settings, this temporal gap will result in the last syllable of a relatively prosodically salient unit being assigned a relatively long duration.[1] Thus, in 1.1 it will be noticed the last syllable of "Buffalo" and "cat" are assigned longer durations (a quarter note, or two x's on the smallest metrical level) compared to the syllables "Buf," "fal," "in" and "the" which do not mark junctures of prosodic boundaries. In 1.2 a grouping/prosodic hierarchy mismatch is created by a relatively long duration being assigned within the clitic group "in the hat" and within the word "Buffalo."

1.1 and 1.2 provide evidence that the informal principle for constituency matching we are suggesting is operative. However, the evidence is only circumstantial in that the two sequences "Buffalo Bill" and "Cat in the hat" while stressed identically differ not only in their prosodic and syntactic structure but in much else, most notably in their local phonological structure. More conclusive evidence would be provided if it could be shown that two sentences differing only in their prosodic/syntactic structure require distinct settings. Examples of such sequences which control for all linguistic elements other than prosodic constituency are known to word mavens and linguists as "oronyms." Examples include

1.3 Oronyms

a) We needed a cantor / We need a decanter.
b) Sea defenses / Cedar fences.

[1] Duration should be understood in the sense of temporal distance between attack points of events. See section 4 for discussion.

c) Doggy dog world / Dog eat dog world.
d) Almost interesting / All most interesting.

It will be seen that certain rhythmic patterns can be associated with 1.3 a-d which appropriately match textual and metrical stress but which either succeed or fail to create appropriate matchings between the relevant constituent units. When the correspondence of prosodic and grouping structure is "in phase" settings tend to be relatively acceptable:

1.4. a)

[we need ed][a can tor]

b)

[we need][a de can ter]

Those which are out of phase are relatively problematic.

1.5 a)

[we need ed][a can tor]

b)

[we need][a de can ter]

From the standpoint of constituency structure, the two linguistic sequences are identical except in one respect: a prosodic boundary intercedes between the third and fourth syllables in 1.4 a and 1.5a and between the second and third syllables in 1.4 b and 1.5 b.

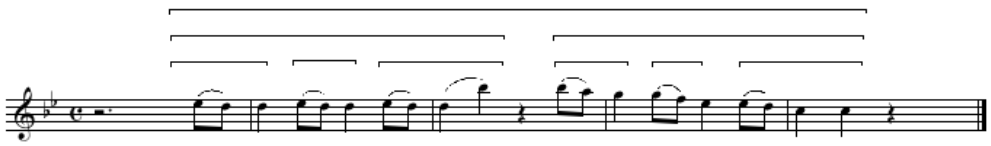
Settings of the remaining oronyms in 1.3 will be seen to provide further verification of the principle proposed above. Also providing verification is an informal study I conducted among ten colleagues: given a choice between in phase and out of phase matchings, there was unanimous agreement that the in phase matchings were preferable.

2. Musical Grouping

A musical phrase, motive, section or movement are all instances of musical grouping in that each is heard as a discrete musical unit bounded off from other constituent units on either side. The difference between these units is in the relative size, or, more precisely, in the level of the grouping hierarchy that each implicates. Grouping will be seen to be hierarchical in that smaller constituent sequences combine together at higher levels of grouping structure with other groups to form larger units up to and including the entire piece.

Example 2.1 indicates a typical grouping structure, that which listeners assign to the opening of Mozart's Symphony no. 40:

2.1.



The lowest level of the hierarchy indicated in 2.1 contains six groups composed of three or four notes. These combine into two intermediate level groups of ten notes each themselves combining to form the largest group consisting of all twenty notes.

The constitution of these groups is inferred by listeners based on cues provided by characteristics of pitch, timbre, density, articulation and timing of the events and their relationship to the underlying harmonic structure of the work. But while all of these characteristics can be significant, as just mentioned, probably the most definitive cue for a grouping boundary results from the manipulation of temporal proximity: a relatively long temporal distance between two successive events, or more specifically, between the temporal locations of the attack points of each event tends to be a fairly decisive indication of a grouping boundary. When this temporal distance is relatively, several seconds or more, the musical juncture is interpreted as defining a major sectional division within a piece, possibly a distinct movement or section. When the temporal distances between events are relatively short as they are in 2.1, those which are relatively long compared to those in the immediate vicinity map out low levels of the grouping hierarchy, informally described as a motive, cell, or perhaps a phrase.

This single factor-temporal proximity- will be seen to determine the grouping structure applied to 2.1. All grouping boundaries correspond to relatively long gaps in the sense just described with the most prominent grouping boundaries-those which conclude the entire group-corresponding to the two largest temporal gaps. This principle can be defined with a fair degree of precision as it is in, for example, Lerdahl and Jackendoff (1983). For our purposes here, it will be useful to define the proximity principle less rigorously:

2.2

Proximity principle: assign a grouping boundary to the right of a musical event e_n if the temporal distance between the attack point of e_n and that of

the following event e_{n+1} is greater than those of other pairs of adjacent events in the immediate vicinity.

It will be seen that the temporal duration between the attack points associated with the events 3 and 4 in the Mozart excerpt in 2.1 is greater than those of the pairs 2 and 3 and 4 and 5. The proximity principle will assign grouping boundaries in the appropriate locations both at this juncture and the other locations indicated by the horizontal brackets in 2.1.

To reiterate, the proximity principle just identified is by no mean the only cue according to which listeners infer the presence of grouping boundaries. For example, sharp changes in patterns of articulation, dynamic, pitch contour, or instrumentation will create grouping boundaries at the location where the change occurs. For our purposes it will be sufficient to focus solely on temporal proximity for it will be seen to be sufficient to identify those grouping boundaries relevant to the cases we will concern ourselves with. Furthermore, since it does not implicate pitch, harmony or instrumental tone color, it is able to assign grouping boundaries in the unpitched idioms (e.g. chants, poetic recitations and normal speech) in addition to conventional pitched vocal music.

Most importantly, it, and it alone, will be seen to be sufficient to predict the correct results: the proximity principle, in assigning a grouping boundary after relatively long events, creates an in phase relationship between grouping structure and prosodic structure in those settings we have dealt with thus far.

2.3 a)

[[we need ed] [a can tor]]

b)

[[we need] [a de can ter]]

In each setting, the proximity principle assigns a grouping boundary after the longest event. The same principle holds for other settings discussed above. The prosodic boundaries in the settings "Buffalo Bill" and "Cat in the hat" will also seen to be made in phase with the grouping hierarchy derived by simple application of the proximity principle. This relationship is ensured by "lo" and "cat" being assigned relatively long durations which create grouping boundaries at the appropriate locations. Those rhythmic settings of oronyms which the poll respondents indicated as preferred will also be seen to contain grouping boundaries predicted by proximity principle at the correct locations.

3. Higher level constituency matching

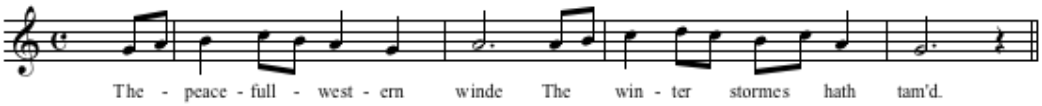
Thus far we have been concerned with relationships between relatively low levels of the grouping and prosodic hierarchies specifically the relationship between the word and low level melodic groups-those which would be referred to as melodic cells or motifs. It will be seen that constituency matching also applies to higher levels of the grouping and prosodic hierarchies as well. The following will be concerned with evidence for such correspondences of two different sorts. To begin with, we observe that

those with relatively developed musical intuitions tend to find natural settings which manifest in-phase constituency matching at intermediate and higher levels of the grouping and prosodic hierarchy.

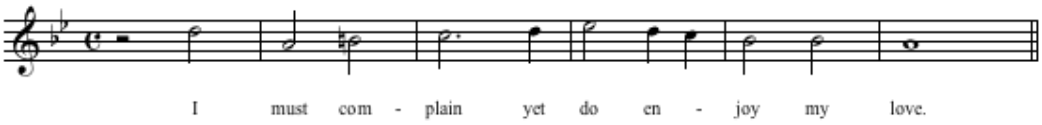
A good indication of these mechanisms at work at an intermediate level can be seen in the following settings by Campion.

3.1

a) Thomas Campion: “The Peacefull western winde” (1601)



b) Thomas Campion: “I must complain” (1601)



Grouping boundaries imposed by the proximity principle occur after the long events assigned to the syllables "winde," "tam’d," "(com)plain," and "love." Thus, grouping structure (predicted by the proximity principle) is in phase with the prosodic hierarchy. But whereas the relevant prosodic units in the previous examples were at the word or clitic group level, here the relevant prosodic constituents are at higher levels of the prosodic hierarchy-namely at the level of the phonological phrase:

3.2
a) [[[The Peaceful western winde]] [[The winter stormes] [hath tam’d]]]

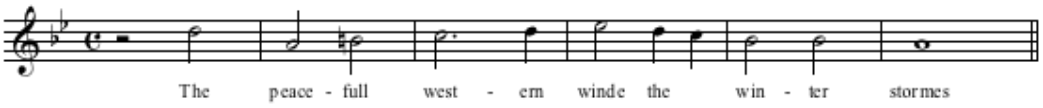
b) [[I must complain] [yet do enjoy] [my love]]

That Campion chose to match constituents by assigning relatively long events to the last unit of each prosodic constituent is, as we shall see, typical of English renaissance

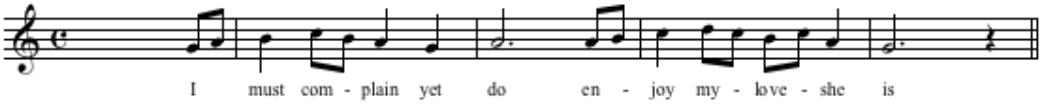
text setting practice. But more significantly, the matching of constituents appears to be consistent with normal intuitions with respect to what those with musical experience regard as the most natural arrangement of text and tune. That this is the case can be seen when the texts are interchanged such that a grouping/prosodic constituency mismatch is created:

3.3

a)



b)

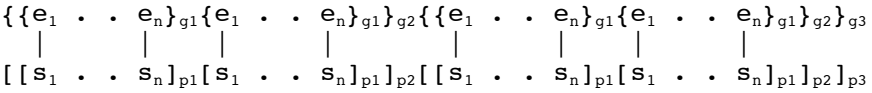


That both texts are iambic insures that textual and metrical accents will be properly aligned in the constructed settings. However the redistribution of poetic feet results in mismatches between grouping structure and prosodic constituents. The intuitions implicated here are, admittedly, somewhat more subtle than in the previous cases. That they are at least to some extent operative is indicated by the results from the respondents of the poll previously referred to. When asked to try to distinguish between constructs and originals most were able to detect the original settings. When asked to identify the locations which seemed most unnatural most identified the dotted half note assigned to “do” as significantly less natural than the same event assigned to “winde”.

4. Consituency matching in practice: Dowland Lute Songs

The responses of this informal poll will be seen to be consistent not just with the practice of Campion in these songs but appear to be mirrored more generally in the practice of composers in the English renaissance vocal literature of which the lute songs of John Dowland are perhaps the best known representatives. Before we discuss these, it is worth mentioning that if it is the case that matchings of the sort we have been describing are relatively natural, it is easy to formally represent the optimal relationship between text and tune which informs both the intuitions of musically experienced listeners and, possibly, composers' practices along these lines.

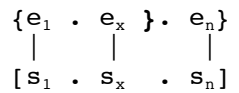
4.1.
Idealized constituency matching



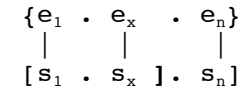
In 4.1, e refers to musical events which are assigned to syllables, s. Musical groups derived from sequences of events represented here by braces {} are co-extensive with prosodic boundaries represented by brackets [].

What tends to be avoided are constituency mismatches. These are of two types: when a grouping boundary intercedes between two syllables which are not separated by a prosodic boundary (4.2a) or when a prosodic boundary occurs within a musical group (4.2b).

4.2 Constituency Mismatches
a)

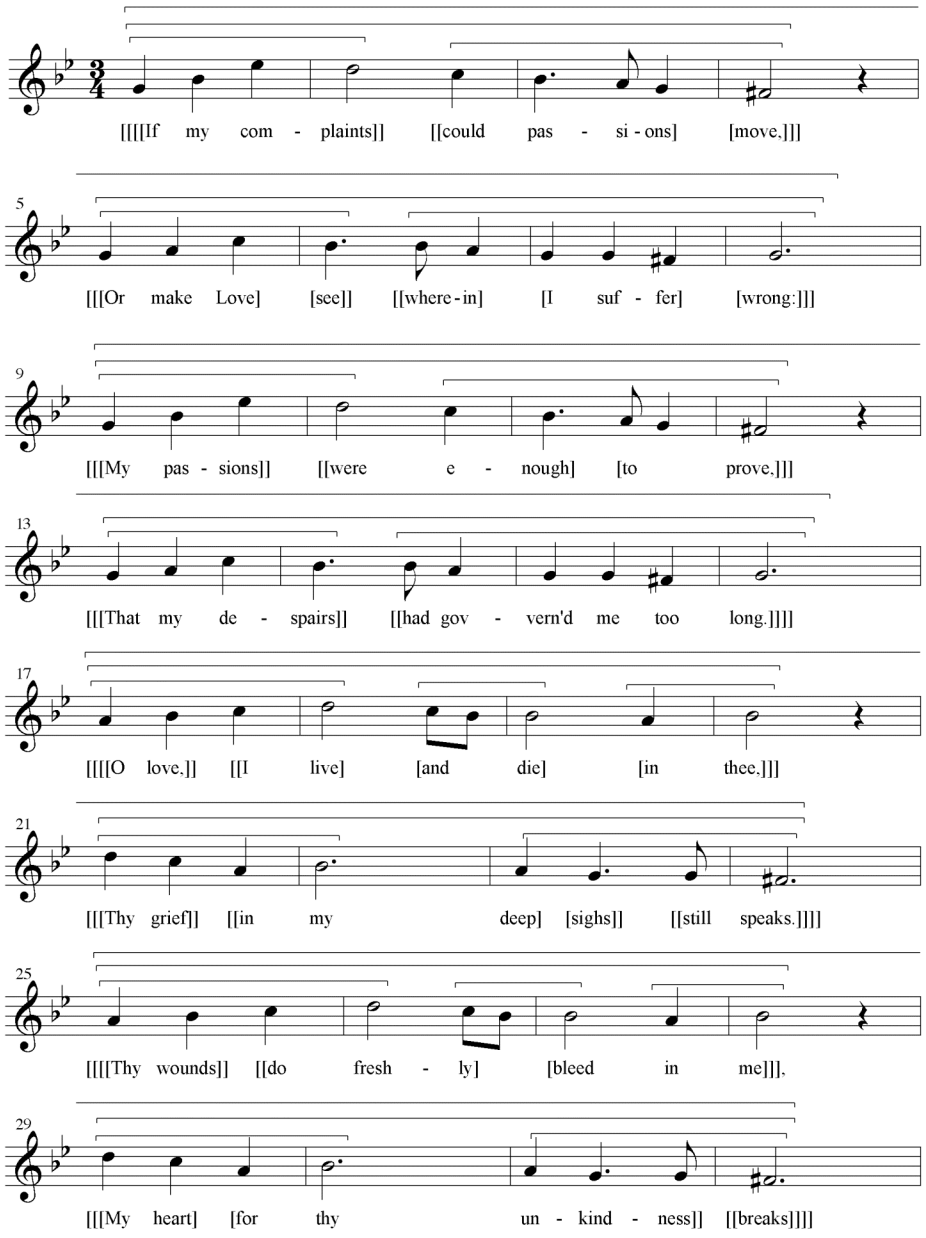


b)



As mentioned previously, within the renaissance idiom, grouping boundaries tend to rely heavily on the proximity principle for grouping closure. In phase constituency relationships between text and tune will, therefore, result in final syllables of prosodic constituents being assigned to relatively long events. We would, of course, not expect settings to be maximally in phase with no constituency mismatches whatsoever, however something close to it can be seen in Dowland's "If my complaints":

4.3 John Dowland “If my complaints” (1597)



The verification of constituency matching requires two separate analyses, one linguistic indicating the prosodic hierarchy, the other musical indicating grouping structure. For the former we propose four levels of the prosodic hierarchy represented by brackets in 4.3. These include the utterance (coextensive with the sentence), the intonational phrase (corresponding roughly to major phrasal units), and the phonological phrase, (corresponding to minor phrasal units). We represent these by four, three and two brackets respectively. The lowest level of the prosodic hierarchy represented here, the clitic group, consists of one content word representing a major lexical category (a noun, verb or adjective) adjoined leftward or rightward to words from minor lexical categories (articles, pronouns, prepositions etc.).[2]

The following chart itemizes the representation of these categories:

4.4 Occurrence of prosodic boundaries

Prosodic Boundary type	No of brackets	No. of instances.	% of total
Utterance	4	5	4
Intonational phrase	3	7	6
Phonological phrase	2	14	13
Clitic Group	1	14	13
No salient prosodic boundary	0	64	63
Total syllables		104	

As before, we will take the presence of relatively long events to be our primary diagnostic for grouping boundaries. As explained previously, "long" events are characterized by relatively large "inter-onset intervals." IOIs (as they are referred to in the music-perceptual literature) are defined as the time interval between the attack points of two successive events. IOIs are of five types in this song which, as can be seen in 4.4,

[2] We recognize that the formal derivation of the prosodic structure remains controversial among linguists.

extend from an eighth note, to the most common notated event the quarter note, the dotted quarter, the half and the dotted half (or half plus quarter note rest in some instances). We will refer to this latter event as maximally long, or L.

4.5

IOI (in eighth notes)	No. of occurrences	% of total
6 (dotted half) (L)	14	13
4 (half)	8	7
3 (dotted quarter)	8	7
2 (quarter)	64	60
1 (eighth)	12	11
Total	106	

A comparison of 4.3 and 4.4 allows us to see more clearly the relationship between the prosodic hierarchy and syllabic length. All instances of the two most highly ranked junctures within the prosodic hierarchy-the utterance and the intonational phrase are assigned by Dowland to the most pronounced type of grouping boundary, that created by what we refer to a the maximally long event L.

4.6

Prosodic boundary	No. of occurrences Assigned to L	% of total assigned to L
Utterance	5	100
Intonational Phrase	3	100
Phonological Phrase	0	0
Clitic group	0	0
No boundary	2	3

Given that maximally long events account for 13% of total events, the above distribution implies a strong correlation between prosodic and grouping structure—what we refer as high level constituency matching of the grouping hierarchy.

The prosodic hierarchy represented here is admittedly one of several possible assignments.

At lower levels of the prosodic hierarchy the correspondence between prosodic and grouping hierarchy begins to degrade immediately. In fact, there appears to be a negative correlation in the above distribution. While not necessarily surprising, this result would seem to be inconsistent with the influence of low level constituency marking observed in examples in (2). To investigate this apparent anomaly, it will be useful to define a relatively long event L' as those associated with IOIs in the upper half of the distribution of lengths indicated in 4.3. The following table identifies these.

4.7

Prosodic boundary	No. of occurrences Assigned to L'	% of total assigned to L'
Utterance	5	100
Intonational Phrase	3	100
Phonological Phrase	7	50
Clitic group	4	29
No boundary	8	12

Of 106 events, 33 (32%) are categorized as L'. The assignments indicated in 4.7 appear to suggest a weak correlation between phonological phrase boundaries and relatively long events as well as weak evidence for a negative correlation between absence of prosodic boundaries and relatively long events. There appears to be no evidence for influence of the clitic group level on the text setting practice of Dowland as realized in this song.

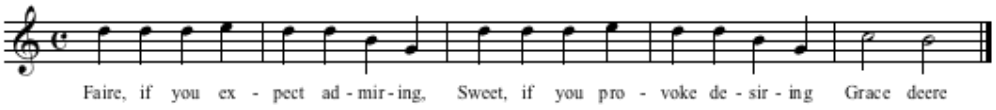
5. Meter and text setting: general discussion

Given that the preceding analysis is limited to one song, it is reasonable to entertain the possibility that it may only reveal idiosyncratic choices on Dowland's part applicable to "If my complaints." This objection can be addressed by a quick perusal of collections such as Greenberg et. al. (1955). These songs will invariably contain a few

constituency mismatches, however the normative structure will be found to be that identified in 4.1.

Furthermore, some of these constituency mismatches will be apparent and not real in that they will result from the failure of the simplified variant of grouping structure to detect grouping boundaries when they result from the manipulation of musical parameters other than length. As mentioned, among these parameters are pitch and harmony. For example, the proximity principle would fail to detect grouping boundaries in the following passage from Campion:

5.1 Thomas Campion: “Faire if you expect admiring” (1601)



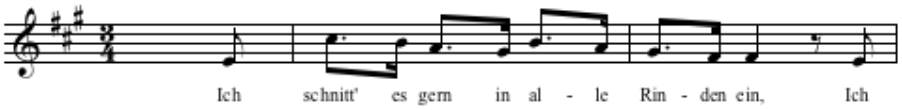
A more comprehensive analysis of 5.1 would assign a grouping boundary based not just on temporal proximity but also similarity. In this case the exact repetition of the sequence of the first eight events suggests a grouping boundary where the repeated unit is reinitiated, that is, after the eighth quarter note.[3] This grouping boundary coincides with the prosodic boundary following "admiring" and as a result, grouping and prosodic structure turn out to be in phase in 5.1.

Even in styles which are highly reliant on pitch and harmony for the perception of musical constituency, the proximity principle probably remains the primary cue for the perception of grouping structure. Among these idioms is that of Schubert whose iambic pentameter settings are studied in detail by Fehn and Hallmark (1983). It will be noticed

[3] An additional cue for a grouping boundary is provided by pitch structure, namely the upward leap of a perfect fifth at this point.

that what Fehn and Hallmark identify as type Y results in most cases from a grouping boundary imposed by the proximity principle coextensive with the end of the poetic line.

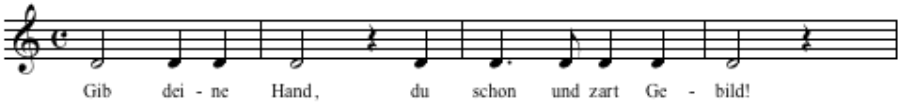
5.2
Schubert: Ungeduld



Fehn and Hallmark's X type results from an expansion of durations assigned to the final syllable of the highest ranked prosodic unit within the line. This usually occurs after the second or third metrical foot. In so doing, Schubert creates a line medial caesura similar to that which is assumed for syllabic verse types, most notably the Alexandrine.[4]

5.3

Schubert: Gib deine hand



In general, the line end corresponds to a prosodic boundary, thus constituency matching is achieved in most Schubert settings of iambic pentameter texts. The exception, of course, is in cases of enjambment. Here, the composer is presented with a choice between responding to the poets versification manifested in the imposition of a line boundary and the prosodic structure inherent in the utterance. The former will create a

[4] Francois Dell points out (P.C) Schubert's grouping boundaries do not occur at the locations associated with those legislated for cesurae by French prosody.

constituency mismatch while the latter will, arguably, result in disregarding the poet's intentions. How this conflict is resolved is a topic we will leave aside for further study.

6. Normative metrical types within Metrically Rigid Verse

As mentioned earlier, because the proximity principle is not reliant on pitch, it is able to assign grouping structure to pitchless idioms most notably the poetic tradition described by Oerhle (1989) as "metrically rigid verse." These verse types are defined by a more or less unambiguous pattern of rhythmic declamation. Those fluent in the relevant metrical idiom are able to decode from lines of text inherent underlying rhythmic patterns. Often, in fact, the lines are lyrics to pre-existing songs or chants whose basic rhythmic pattern is known in advance. Acceptable lines of metrically rigid verse are determined by whether they can be naturally declaimed within what is understood to be a pre-existing rhythmic context. Examples provided by Stein and Gil (1980) and Burling (1970) demonstrate that this tradition exists within many linguistic and musical traditions.

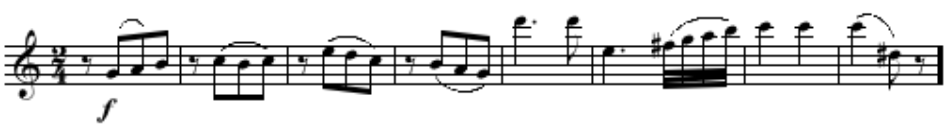
One of the characteristics of metrically rigid verse which cannot be incorporated within either traditional prosody or within variants of generative metrics is the possibility of empty metrical positions- positions identified in the underlying meter but which are not occupied by a syllable. Examples of these empty positions are common in nursery rhymes, but also art verse, most notably the "fourteener" discussed in detail by Attridge (1982).

It has been noted both by Attridge and by Hayes and MacEachern (1998) (hereafter H+M) that the most likely location for an unoccupied strong position is the fourth beat. In fact, it appears that this is the only strong position which can be

unoccupied in verse patterns; all others (the first, second and third) are required to be occupied. This is indicative of an essential difference between music and verse: while patterns having "downbeat" rests such as 6.1 are common within both vocal and instrumental music, creating a somewhat breathless affect, there are no poetic verse types which leave the first metrical position unoccupied in this way.

6.1 Mozart's K. 425 Linz Symphony, fourth movement, mm. 83-87.

a)



b) Grid notation. (S refers to positions occupied by musical events)

```
      S  S  S     S  S  S
      S  S  S  S  S  S  S
   S        S  S        S
   S     S     S     S
||:X  X  X  X  X  X  X  X :||
   X     X     X     X
```

Furthermore, as H+M note, certain distributions of "3" and "4" type lines (the latter occupying the fourth position, the former omitting it) are disproportionately represented within the corpus. More importantly for our purposes, these "feel" much more natural than others. Among these is that which H+M categorize as the quatrain 3343. The limerick which has an underlying ternary subdivision of strong beats is an example of this type:

6.2

```
      S S S S S S S S S          3
      S S S S S S S S S          3
      S S S S S S S S S S S S    4
      S S S S S S S S S          3
||:   X X X X X X X X X X X X     :||
        X     X     X     X
```

A variant of this pattern is the nursery rhyme "Hickory Dickory Dock."[5]

6.3

```
        S S S S S S S              3
    S S   S S   S S S              3
    S S   S S   S S   S S          4
      S S S S S S S                3
||: X X X X X X X X X X X X       :||
      X     X     X     X
```

A binary instance of the 3343 type is the nursery rhyme “Peas porridge hot.”[6]

6.4

```
      S   S S S            3
      S   S S S            3
      S   S S S S S        4
      S   S   S            3
||:   X X X X X X X X :||
      X   X   X   X
```

Also common is the 4343 type found in chants and simple songs such Bah Bah Black Sheep.[7]

6.5

```
      S   S   S   S        4
      S S S S S            3
      S   S   S   S        4
      S   S   S            3
||:   X X X X X X X X :||
      X   X   X   X
```

[5] The lyrics associated with these most familiar of nusery rhymes are included here for non-native speakers. Hickory, dickory dock/the mouse ran up the clock/the clock struck one/and down he run//hickory dickory dock.

[6] Peas porridge hot./ Peas porridge cold./ Peas porridge in the pot/ Nine days old.

[7] Bah, bah black sheep/ Have you any wool/ Yes sir, yes sir/ Four bags full.

The 4343 type also occurs in ternary form in the nursery rhyme "Old Mother Hubbard"[8]

6.6

```
     S  S  S  S  S     S  S  S  S  S        4
  S  S  S  S  S     S  S                    3
  S  S  S  S  S     S  S  S  S  S           4
  S  S  S  S  S  S  S  S                    3
‖: X  X  X  X  X  X  X  X  X  X  X  X  :‖
     X        X        X        X
```

H+M note that certain quatrain types are both extremely uncommon and sound odd. Among the more jarring of these is the form 3344 for which H+M provides a construct[9] which substitutes “Hickory, dickory, dickory, dock” for the final line of 6.3 resulting in

6.7

```
     S  S  S  S  S  S  S                    3
  S  S     S  S     S  S                    3
  S  S     S  S     S  S     S  S           4
     S  S  S  S  S  S  S  S  S  S           4
‖:  X  X  X  X  X  X  X  X  X  X  X  X     :‖
     X        X        X        X
```

Our discussion of constituency matching will be seen to provide some insight on the intuitions which enter into these judgements. Before we demonstrate this, it is worth noting what a correct analysis of well-formed quatrains should look like. For the answer to this we appeal to H+M's terminology which identifies the quatrain and to a lesser degree couplet divisions as "cadential" or closed. This implies a hierarchy of grouping which mandates the strongest closure occurring at the completion of the quatrain, a lesser

[8] Old Mother Hubbard went to the cupboard,/To give her poor dog a bone./ But when she got there, her cupboard was bare;/ And so the poor dog had none.

[9] Hickory, dickory dock/the mouse ran up the clock/the clock struck one/and down he run//hickory dickory dickory dock.

degree of closure following the couplet and only weak closure following the first and third lines. Thus, something like the following will be expected:

6.8
a) $L_1]L_2]]L_3]L_4]]]$ b) $L_1L_2]L_3L_4]]$ or equivalently c) $L_1]C]]L_3]Q]]]$ d) $L_1C]L_3Q]]$

Stated slightly differently, preferred quatrains will be expected to adhere to the relationship

6.9 $|Q| > |C| > |L_1| = |L_3|$

where |X| is to be read as indicating the number of brackets assigned to the right edge of a quatrain element X.

Relatively unnatural settings will be expected to contain relatively strong grouping boundaries at positions where "closure" is least expected and vice versa. Thus, what we will call unpreferred quatrains will be likely to meet at least one of the following conditions:

6.10 A quatrain L_1CL_3Q is unpreferred if

a) $|L| > |Q|$ b) $|L| > |C|$ c) $|C| > |Q|$ d) $|L_1| <> |L_3|$

7. Analysis and prediction of preferred quatrain types.

Our objective here will be to provide a constituency analysis which unproblematically assigns a grouping structure to each of the sixteen possible quatrain types-those which result from all combinations of 3 and 4 type lines. If the assigned analysis is correct, those quatrains which are heard as most natural should at minimum a) not embody any of the objectionable characteristics specified in 6.10 and, at best, b) meet the conditions of the optimal quatrain 6.9. Conversely, those which are heard as deviant should a) (obviously) not be designated by the constituency analysis as preferred

according to 6.9 and b) should be expected to embody at least some of the violations itemized in 6.10.

The first step of the constituency analysis of the quatrain types is to apply the proximity principle. Doing so will result, by the definition 2.2, in assigning a grouping boundary after relatively long durations, namely at the end of those lines whose text extends only to the third strong beat, what we will call 3-type lines. The final syllable of 4 type lines, those which occupy the fourth strong position extends at most only two durational units to the first strong beat and therefore do not allow for the assignment of a grouping boundary by proximity. Applying this interpretation yields the grouping structures to the lines in question:

7.1 (letters indicate violations of criteria in 6.10)
a) 3]3]43] d
b) 43]43]
c) 3]3]44 a, c, d

The analysis fails on three grounds: first it fails to assign an optimal grouping structure to the most natural forms 7.1 a) and b), and more seriously, 7.1 a) is designated as unpreferred. As a guide to how we might alter our analysis we will also note that the constituents represented in 7.1 a) and c) are problematic in their excessive segmentation of the quatrain. As noted in GTTM, relatively small grouped units-in our case a single line-tend to be avoided. Consistent with this observation, it will be seen to be useful to apply the proximity principle more strictly, assigning grouping boundaries to units no smaller than pairs of lines. The proximity principle for quatrains can be formalized as follows:

> 7.2 Proximity (quatrain version): In any pair of successive lines, if the syllable count associated with the first is greater than that associated with the second, insert a bracket at the end of the second.

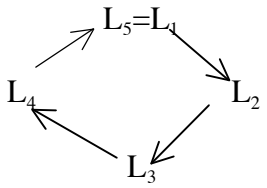
It will be noticed that the requirement that the second line of a pair be more closed than the first implies that only one type of adjacent pair will be assigned a boundary, namely the 43 pair. Applying this analysis to the lines in question yields:

7.3
a) 3343]
b) 43]43]
c) 3344

With this alteration, none of the assigned structures meets any of the conditions specified in 6.9 or 6.10. All of the lines 7.3 a-c are designated as marginal, as none is as classified as preferred or unpreferred.

The next step is to add a necessary additional element into the proximity analysis: that is, to recognize the cyclic structure of these forms indicated by the repeat signs above. Generally, these forms contain a several quatrains strung together such that the final line of the quatrain is followed immediately by the onset of the next.

7.4



Since the fourth line now connects to the to the first line of the quatrain the proximity principle will apply at this juncture. When the proximity condition is met, i.e. when a 43 pair "wraps around" from the last to the first line of the of the form, a grouping boundary is assigned after the "short" first line. The revised proximity principle predicts that grouping closure will be heard after the wrapped around first line of the 3344 quatrain.

Thus it is assigned the structure 3]344 and is designated as unpreferred according to 6.10, correctly predicting that this quatrain will be heard as somewhat awkward.

The revised proximity principle does not apply to 3343 and 4343 quatrains and these remain incorrectly designated as marginal. To account for these forms, we need to bring to bear a second principle, namely that of similarity alluded to earlier. Similarity has at its basis the intuition that identical units in close proximity tend to be grouped together. For our purposes, this will mean that the presence of a sequence of two similar metrical units (whether lines or couplets) mandates the imposition of a grouping boundary at the end of the sequence.

We can formalize the similarity principle as follows:

7.5 Similarity: If two adjacent lines or couplets have the same syllable count, insert a bracket after the second line or couplet.

Applying the similarity principle to the three quatrains under consideration yields

7.6 a) 33]43 b) 4343] c) 33]44]].

Combining both similarity and proximity yields the following analyses:

7.7 a) 33]43], b) 43]43]] c) 3]3]44]].

One problem remains: in 7.7 a): the couplet boundary receives equal closure compared to the quatrain boundary and thereby fails to meet the optimal conditions specified in 6.10. This can be corrected by positing that boundaries imposed by proximity achieve a higher degree of closure than those assigned by similarity. We will designate this distinction by mandating that grouping boundaries assigned by proximity receive two brackets. The revised proximity principle reads as follows:

7.8 Proximity for quatrain (revised): In any pair of successive lines, if the syllable count associated with the first is greater than that associated with the second, insert two brackets at the end of the second.

In addition to yielding the correct results, to posit this revision is also intuitively reasonable from a musical standpoint: an actual gap a pattern of attacks is likely to result in a more decisive perception of closure than the perception of similarity. More significantly, the revised proximity principle now designates both natural sounding and most commonly attested quatrains preferred status and designates the deviant quatrain as unpreferred according to criteria 6.10 b) and c).

7.9 a) 33]43]] b) 43]]43]]] c) 3]]3]44]] *

8. Derivation of attested quatrain forms.

The previous discussion might appear to involve an overly complicated theoretical machinery to correctly analyze only three forms. However, now that we have formulated what appear to be some of the underlying principles, we can test the predictive power of the two rules by applying them to the 16 quatrain types resulting from the distribution of 3 and 4 type lines.

8.1 All Quatrains (violations of 6.10 noted in italics

a) 4444 > 44]4]4]] *d*
b) 4443 > 44]4]3]] *d*
c) 4434 > 44]3]]4 *a,b,c,d*
d) 4344 > 43]]44] *c*
e) 3444 > 3]]44]4] *a, b, d*
f) 4433 > 44]3]]3]] *b, d*
g) 4343 > 43]]43]]] *no violations*
h) 4334 > 43]]3]4 *a, c, d*
i) 3443 > 344]3]] *b, d*
j) 3344 > 3]]3]44]] *b, d*
k) 3434 > 3]]43]]4] *a, b*
l) 4333 > 43]]3]3] *b, c, d*
m) 3433 > 343]]3] *a, b, d*
n) 3343 > 33]43]] *no violations*
o) 3334 > 3]]3]3]4 *a, b, c, d*
p) 3333 > 33]3]3]] *d*

The grouping structures assigned by the combination of proximity and similarity meet the criteria designated above: they assign preferred structure to two commonly attested types, 8.1 g) and n) 3343. They also mark as unpreferred all other forms, though some are less unpreferred than others by virtue of their violating fewer of the conditions in 6.10. Several of the most commonly attested quatrains violate only one criteria. And highly violative quatrains-those meeting two or more of the conditions in 6.10 do appear to be, in most cases, awkward and are uncommon. These results offer partial confirmation of the relevance of constituency analysis to metrical form.

A more detailed study of these lines would not be productive here since our intention is not to demonstrate that constituency is the sole factor in determining acceptable correspondences of text and tune but rather to demonstrate how much can be explained by a constituency analysis independent of metrical structure -i.e. an analysis which ignores matching of textual to metrical accent and is based entirely on relative durations and the groupings which these patterns imply. Furthermore, it should be obvious that many salient aspects of lines are not accounted for within the schematic approach we have provisionally adopted. The reduction of the data into two line types, 3-type and 4-type characterizes only the final events of each grouped unit. As such, it omits from the analysis what happens everywhere else, most notably, at the beginning of the line. [10]

[10] Since H+M's analysis also only considers the final two metrical positions, it is not able to distinguish between lines which differ in structure at the beginning of the line. That these are significant in listeners' judgment of similarity is both intuitively obvious and turns out to be necessary for the correct analysis of these forms.

An indication that incorrect predictions can result from this overly reductive categorization of line types can be seen in 8.2. All of the lines are taken as generic 4 type lines and the grouping analysis 44]4]4]] assigned in 8.1a) is derived.

8.2 (Humpty-Dumpty)

```
        S    S S    S S S S S              4
        S    S S    S S S S S              4
        S S  S S  S S S S S S S             4'
        S S  S S  S S S S S S S             4'
||:     X X  X X  X X X X X X X  X X :||
        X      X      X      X
```

In fact, the relatively long durations assigned to the first and third syllables in the first and second lines, while too short and too proximate to require a grouping boundary assigned by the proximity principle delineate similarity with a clear differentiation marked by the appearance of an event on the second metrical position in the third line. The similarity analysis should not be applied to the line type 4444 but rather to the line type 444'4'. Doing so yields the grouping structure 44]4'4']]. This analysis now meets the optimal conditions for the quatrain hierarchy specified above in 6.9, a designation which is probably appropriate given that it both common and is easily internalized. We will not attempt to revise the similarity principle accordingly here since it is not our intention to claim that grouping and constituency matching can explain everything. Rather, our objective, both in this section elsewhere, is to argue that it appears to explain a great deal.

9.0 Conclusion

The above discussion is focussed primarily on what we have called constituency matching: the relationship between musical groups and prosodic constituents. We do not mean to suggest that all poetic texts necessarily should be viewed as having musical or

even rhythmic characteristics. In certain poetic meters, syllables may be chosen based on constraints which may have little to do with the temporal characteristics which are required for the perception of rhythm or grouping. Among these metrical traditions are pattern poems. One of these, according to M. Halle (1987) is Psalm 137 which, according to Halle when properly scanned forms an image of the first temple in Jerusalem. Obviously, any attempt to assign grid structures to these sequences of syllables would be unproductive as would an attempt to assign musical constituency to them. The reason is that the arrangement of syllables was dictated by the requirements appropriate to creating a pattern in visual space, not a pattern in audible time.

These sorts of poetic meters we believe are the exceptions. The norm for texts arranged in lines is to imply that syllables create patterns in time.[11] Generally this means one of two things: either the texts are actual song or chant lyrics to be assigned to a pre-existing tunes or rhythmic patterns. Or they are improvised chants or songs whose temporal arrangement requires exercising the creative faculties of those engaging the texts. (Halle, 1997, Halle and Lerdahl, 1993). In either case, intuitively based constraints on matching texts to the grouping and metrical structure of the "tunes" (either pitched or unpitched) is what determines the structure of acceptable lines. While this study is specifically oriented towards identifying one characteristic of this cognitive faculty, namely, constituency matching, we hope that it will be taken as further evidence that the perceived quasi-musical relationships inherent in texts are in many cases a crucial component in determining what sorts of linguistic sequences can be construed as lines, whether these are categorized as poetry or song lyrics.

[11] See Sapir (1921) for an explicit statement of this position.

Acknowledgments:

Thanks to Francois Dell and Nigel Fabb for detailed criticisms on a draft version of this paper and to the participants in the "Words and Music" conference for many helpful suggestions.

References

Attridge, D. (1982). *The Rhythms of English Poetry*. London: Longman.

Burling, R. (1970). *Man's many voices; language in its cultural context*. New York: Holt, Rinehart and Winston.

Campion, T. (1601) *The Songs from Rosseter's Book of Airs*. Edited by Edmund H. Fellows. In Thurston Dart (1969) *The English Lute-Songs*. Series I, Vol. 4/13. London: Stainer and Bell, Ltd.

Dowland, J (1597) *The First Book of Ayres* Edited by Edmund H. Fellows. In Thurston Dart (1969) *The English Lute-Songs*. Series I, Vol. 1/2. London: Stainer and Bell, Ltd.

Dowland, J (1600) *Second Book of Songs* Edited by Edmund H. Fellows. In Thurston Dart (1969) *The English Lute-Songs* Series I, Vol. 5/6. London: Stainer and Bell, Ltd.

Fehn, A.C. and R. Hallmark (1983). Text and Music in Schubert's Pentameter Lieder: A Consideration of Declamation. In *Music and Language*. Studies in the History of Music, 1. New York: Broude Brothers.

Greenberg, N., W.H. Auden and C. Kallman eds, (1955). *An Anthology of Elizabethan lute songs, madrigals, and rounds*: New York: Norton.

M. Halle (1987). A Biblical Pattern Poem. In N. Fabb, D. Attridge, A. Durant and C. MacCabe eds. *The Linguistics of Writing*. New York: Methuen.

Halle, J. and F. Lerdahl, (1993). A Generative Textsetting Model. *Current Musicology,* 55, 3-23.

Halle, J. (1997). *A Grammar of Improvised Textsetting*. Doctoral Dissertation New York: Columbia University.

Hayes, B. and M. MacEachern, (1998). Quatrain Form in English Folk Verse. *Language* 64, 473-507.

Hayes, B (1989). The Prosodic Hierarchy in Meter. In P. Kiparsky and G. Youmans, eds., *Phonetics and Phonology, vol. 1, Rhythm and Meter*. San Diego: Academic Press.

Jackendoff, R. and F. Lerdahl (1983). *A Generative Theory of Tonal Music*. Cambridge: MIT Press.

Oerhle, R. (1989). Temporal Structures in Verse Design. In P. Kiparsky and G. Youmans, eds., *Phonetics and Phonology, vol. 1, Rhythm and Meter,* San Diego:Academic Press,.

Sapir, Edward (1921). The Musical Foundations of Verse. *Journal of English and Germanic Philology*. 20:213-228.

Stein, D. and D. Gil (1980). Prosodic Structures and Prosodic Markers. *Theoretical Linguistics* 7. 173-240.