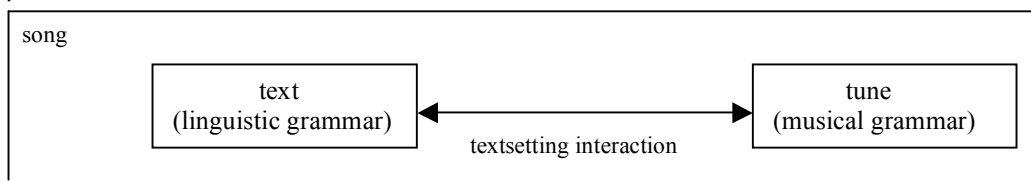


0. Introduction¹

The following discussion takes as its point of departure the perspective advanced in Dell and Halle (2005) which identifies a song as a composite made up of two independent objects, a text and a tune. This underlying organization is graphically represented in the following diagram:

0.1



0.1 makes apparent that tunes and texts routinely exist independently of each other—as instrumental music, on the one hand, and normal speech on the other—and each must be judged as acceptable within the grammars applicable to the respective cognitive domains. In the event that a tune and text are forced into an interactive relationship, as occurs in vocal music, the elements of language—words, syllables, stresses and intonational

¹ Draft versions of this paper have been reviewed in considerable detail by Morris Halle and Francois Dell to whom the author extends sincere thanks. The author assumes sole responsibility for all remaining errors.

structure are required to be matched more or less appropriately with notes, phrases, pitches and rhythms and other elements of musical structure.

Formal studies of text setting e.g. Hayes and Kaun (1996) , Kiparsky (forthcoming), Boone (1999), Lerdahl (2001), Fehn and Halmark (1983), have focused primarily around interactive structure-relationships of text and tune found in songs, usually within strophic song forms. A fundamental objective of our investigation will be to demonstrate that these interactive accounts are incomplete. An acceptable interaction between music and text, while necessary, is not sufficient to determine a well-formed text setting. We will argue that an independent account is required as well: both text and the tune must be judged acceptable as independent objects in order for the composite to be judged acceptable. More specifically, strophic songs in particular require that acceptable tunes of subsequent verses are heard as being sufficiently similar to the original. We will advance what we will call a similarity metric, a mechanism which formally describes the class of variants which can be construed as acceptable on these grounds. The judgement of acceptability is frequently based, as will be seen, on purely musical form as it is defined in the similarity metric without reference to a text; that is, the judgement is made on independent rather than interactive grounds.

Before turning to outlining the similarity metric and to a general discussion of independent constraints on strophic song, we will briefly review some of the interactive accounts which have been offered in the literature as significant for the purposes of textsetting.

1. Interactive constraints.

1.1 Stress Matching

The most readily apparent and most frequently observed relationship² within text setting is that established between linguistic and musical accent, specifically the arrangement of stressed or unstressed syllables and strong or weak metrical positions. Morgan and Janda (1989) (hereafter M+J) is one of many studies demonstrating that mismatches between stress and meter are judged along a gradient range from natural to unacceptable. In their investigation of the capacities underlying these judgments, M+J ask their experimental subjects, who are both fluent speakers and competent musicians, to assign various eight syllable texts to the first phrase of the song “Frère Jacques.”

1.1.1

x				x				x				x			
x		x		x		x		x		x		x		x	
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
C		D		E		C		C		D		E		C	
Frè		re		Jac		ques,		frè		re		Jac		ques	

Among M+J's data is the following significant result: When asked to assign the English text “My brothers dog keeps on barking” to the “Frère Jacques” tune, none of M+J’s subjects produced the tune in 1.1.2 which would result from maintaining the notes of the existing melody:

1.1.2 *³

x				x				x				x			
x		x		x		x		x		x		x		x	
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
C		D		E		C		C		D		E		C	
My		bro		<i>thers</i>		dog		keeps		on		bark		ing*	

² See, for example, Liberman (1975), Oehrle (1989), Palmer and Kelly (1992).

³ The asterisk will be used to indicate constructs which are judged by those fluent in the relevant linguistic and musical idioms to be unacceptable in some significant respect.

Rather most subjects distorted the melody so as to avoid the mismatched positions indicated in bold face and italics—a stressed syllable assigned to a weak position adjacent to an unstressed syllable assigned to a strong metrical position. The most common resolution of the mismatch effected by M+J's subjects was to insert an anacrusis prior to the first note of the existing song shifting the first four events leftward, thereby assigning two notes to the single syllable "dog":

1.1.3⁴

	x				x				x				x		
	x		x		x		x		x		x		x		x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
<G>	C		D		E		C		D		E		C		
My	bro		thers		dog		- ⁵		keeps		on		bark		ing

M+J go on to observe that some languages allow for mismatches and thus the maintenance of the original melody. For example, when asked to assign the text “El perro de mis hermanos” to the same tune Spanish speaking subjects assign syllables to available notes as follows:

1.1.4

					x				x				x		
	x		x		x		x		x		x		x		x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	C		D		E		C		C		D		E		C
	El		<i>per</i>		ro		de		mis		her		man		os

Spanish speakers, unlike English speakers, do not find the mismatch sufficiently objectionable to require alteration of the original melody. This is a specific instance of a general tendency discussed in Dell and Halle (2005): in certain vocal traditions positional

⁴ It will be noticed that the grid in 1.1.3 has been reoriented such that what had been the final position represented in 1.1.2 is now the initial position. This operation is merely a notational requirement of the linear grid. The final and initial positions of the grid are in fact the same position, an organization made explicit in the radial grid proposed in Halle (1999). On the "theft" of initial grid positions also see Hayes and MacEachern (1996) for discussion.

⁵ The hyphen will be used here to indicate continuations of melismatic syllables.

uniformity-the relation of musical events to the metrical grid found in the original tune- can be maintained in the face of what are defined as stress mismatches. In other text setting traditions, significant deviations from positional uniformity are required in order to achieve a proper matching of syllabic stress and the metrical grid.

1.2 Constituency Matching

While much remains to be said as to what constitutes acceptable stress matching it should be recognized that it is not the only interactive constraint governing text setting. As noted in Halle (forthcoming) the structure of text setting is also constrained by the interaction of musical grouping structure and the salient constituent units of language defined by the prosodic hierarchy (see Hayes 1984, Selkirk 1984, Nespor and Vogel 1986). Text substitution also provides clear evidence that an “in phase” relationship between musical and linguistic constituency tends to result in more natural sounding arrangements compared to out of phase relationships which can sound quite bizarre. This can be seen by inserting alternate texts into the Adler-Ross song “Hernando’s Hideaway.” It will be noticed that any non-enjambed iambic tetrameter text substituting for the lyrics in the original will result in the proper alignment of musical groups (indicated by brackets above the staff) and prosodic boundaries (indicated by |):

1.2.1⁶

Whose woods these are I think I know | his house is in the vi-lage though || he
 would not see me stop-ping here | to watch his woods fill up with snow. |||

If, however, an iambic *pentameter* text is substituted in the same melody, the result is an out of phase constituency relationship where grouping and prosodic boundaries do not overlap.⁷ Whatever the aesthetic virtues of out of phase structures like 1.2.2, they are clearly far more difficult to perform and internalize than 1.2.1.

1.2.2 *

The cur - few tolls the knell of part - ing day | the low - ring
 herd winds slow ly o'er the lea || The plow man home ward plods his wea-ry way. | And
 leaves the world to dark ness and to me. |||

⁶ I am grateful to Jay Keyser for providing me with this example.

⁷ Constituency mismatch appears to constitute a more or less exact analogue to enjambment within metrical poetry. We take this one as of several strong indications of the close connection between text setting and prosody.

1.3 Constraints on Melismatic Structure

In addition to constituency and stress matching, a third class of interactive constraints has been recognized at least since the time of the first formal study of text setting by Gasparo Stoquerus (Veltman, 2001). Among the "necessary rules" devised in Stoquerus's *Treatise on Verbal Music* (1570) are those governing the treatment of melismas—syllables assigned to more than one note.⁸ Extensive melismatic passages are common in Renaissance and Baroque styles:

1.3.1 Handel, "Let God Arise"



Melismas are less common and less florid in folk and popular idioms. When they do occur, they generally contain no more than two or three notes:

1.3.2



Stoquerus was the first to state an obvious but easily ignored principle with respect to the relationship of text and tune, namely that whereas the number of notes in a passage can,

⁸ We recognize that, pace Crocker (2005), the term *melisma* is often reserved to refer to assignments of a single syllable to four or more notes with the term *neuma* applying to shorter passages such as those we have been considering here. For clarity, the term *melisma* will be understood to apply to any sequence in which a single syllable is assigned to more than one note.

1.3.5 *



The repeated pitches require separate attacks to make apparent that the two adjacent, identical pitches are distinct events. This alteration has the effect of forcing a repetition of the syllable to which the note is assigned. Thus, “where are” becomes “where are are” and “sage” will be altered to “say age.” Whether esthetic or linguistic intuitions are implicated in this judgment, the awkwardness inherent in same pitch melismas makes them largely unuseable in musical contexts. Of many thousands of melismatic notes in Bach Cantatas I have found only one pair of repeated pitches indicated by the dashed bracketed below:

1.3.6 J.S. Bach Cantata 12, "Weinen, Klagen, Sorgen, Zagen" (aria)



An additional constraint governing the metrical position of melismas hinted at by Stoquerus in Veltman's reading has become a standard pedagogical rule in species counterpoint (Gauldin 1995). Namely, *the continuation of a melisma may not appear in a metrically stronger position than its onset*. Settings which violate this principle are heard as containing an awkward repetition of the initial syllable. Thus,

As pointed out, strophic song forms are defined by the reiteration of a melody onto which distinct texts (variously referred to as stanzas, verses or strophes) are superimposed. As was noted in Halle and Lerdahl (1993), subsequent verses are frequently assigned to texts having significantly differing syllable counts from that in the initial verse. Since each syllable (following Stoquerus) is required to be associated with at least one musical event, the disparity in syllables from verse to verse forces significant deviations in the musical form of each. Among these is the sea chantey "The Drunken Sailor" discussed in Halle-Lerdahl:

2.1 "The Drunken Sailor"⁹

		x			x			x			x			x			
	x	x		x		x		x		x		x		x		x	
	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	(A)	(A)	A		A	A	A		A	A	A		D	(D)	F	(F)	A
a.			What		shall	we	do		with	a	drunk		en		sail		or
b.			Keel				haul		him		till		he's		sob		er
c.	Scrape	the	hair		off	his	chest		with	a	hoop		ir	on	raz		or

From the point of view of independent musical structure, the discrepancy in syllable count results in a musical phrase extending over the four beat pattern which ranges from ten notes in the original a), to as few as seven notes in b) to thirteen notes in c).

When we examine alternative settings of the same text, it is not the case that unacceptability is always due to the violation of interactive constraints. Take for instance other possible settings of the text in 2.1-b found in 2.2.

⁹ In 2.1 the final and initial positions should be understood as representing the same point in time assuming the radial grid as discussed in footnote 3.

2.2 *

		x			x			x			x			
	x	x		x		x		x		x		x		x
	x	x	x	x	x	x	x	x	x	x	x	x	x	x
		A		A	A	A		A	A	A		D	F	A
a)		Keel		haul	him	till		he's	sob	er		-	-	-
b)		Keel		haul	him	till		he's	sob	er				

2.2. a and b are both unacceptable, but whereas the ill-formedness of 2.2 a has to do with interactive constraints, that of 2.2 b has to do with independent musical form. Consider first 2.2 a. The hyphens, as noted earlier, indicate melismas which assign the final syllable of "sober" to the pitches of the original tune. Since the tune is identical to that of the original in 2.1 a, the deviance of 2.2 a cannot be accounted for on "independent" grounds: if the original tune was acceptable, obviously, an exact duplicate must be as well. Rather, the problem here involves the interaction of tune and text, most conspicuously, the stress mismatch resulting from "sober" being assigned to a weak-strong metrical sequence.¹⁰

In contrast, 2.2 b corrects the stress mismatch in 2.2 a with the result that no apparent mismatches of stress and meter remain. The corrected version 2.2 b also contains no constituency mismatches nor, since there are no melismas, can there be violations of the interactive principles governing the placement of melismas discussed earlier. Thus, from an interactive standpoint, 2.2 b appears to be well formed.

From the standpoint of independent musical structure, however, 2.2 b is problematic. Most conspicuously, the absence of hyphens at the end of the form indicate that the events from the original are deleted in this proposed variant. While the melody lacking

¹⁰ An additional problematic independent factor is the melisma assigned to unstressed final syllable of "sober."

the final three pitches of the original is perhaps in itself plausible as the first phrase of some other tune, it would never be heard as a possible variant of the first phrase of "The Drunken Sailor" no matter what words were assigned to the musical events. The proposed tune is heard as violating the integrity of the strophic form defining "The Drunken Sailor" and the setting is therefore rejected not on the interactive grounds according to which 2.2 a is rejected but on independent musical grounds.

As just mentioned, the deletion of the final three events of the original appear to be responsible for the deviance of 2.2 b. However, these are not the only melodic alterations effected by 2.2. b. "ø" in 2.3 indicates the three deletions from the original in the variant 2.2. b:

2.3

		x				x				x				x	
		x		x		x		x		x		x		x	
		x	x	x	x	x	x	x	x	x	x	x	x	x	x
variant*:		A		A	A	A		A	ø	A		D		ø	ø
original:		A		A	A	A		A	A	A		D		F	A
								a						b	c

2.3 makes apparent that some deletions are unacceptable, but not all. For example, the deletion occurring at the metrical position labeled as "a" does not compromise the integrity of the tune. This must be the case since attested, acceptable settings routinely leave this position unoccupied:

2.4

		x				x				x				x	
		x		x		x		x		x		x		x	
	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
		Wake		him	and	shake		him		till		he's		sob	er.
		A		ø	ø	A		A	ø	A		D		F	A
								a							

It is reasonable to conclude that deletion of the events b and c from the original is what results in the unacceptable distortion of the original observed in 2.2 b.

Conversely, in addition to texts having fewer syllables than the original text requiring deletions of events, certain texts are “abundant,” that is, they contain additional syllables which require the addition of events. For example, the text “Put him in the scuppers with a hose-pipe on him” results in the following variant which includes two additional events indicated by a and b.

2.5

		x				x				x				x		
	x	x		x		x		x		x		x		x		x
original	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
variant		A		A	A	A		A	A	A		D		F		A
		A	A	A	A	A	A	A	A		D		F		A	
		a				b										
		put	him	in	the	scup-	pers	with	a	hose		pipe		on		him

The insertions within these metrical positions result in an acceptable alteration in the original melody.

The preceding discussion allows us to draw certain conclusions with respect to the melodic structures which can be construed as acceptable variants of tunes. First, it is apparent that certain metrical positions allow for either vacancy or occupancy by events. Second, other positions appear to require occupancy by events for the setting to be construed as acceptable. Finally, the unoccupied positions of the original tune indicated in 2.5, (positions 1, 2, 4, 8, 12, 14 and 16) suggest an additional logically possible category: those which are required to remain vacant. In the following we will show that all of these categories must be incorporated as necessary components of what constitutes knowledge of a tune.

3.0 Similarity Metric-Descriptive Requirements

In order to examine these propositions more closely, it will be useful to turn to a simpler strophic song. "The Farmer in the Dell" makes use of a similar four beat metrical form to "The Drunken Sailor," but with strong positions separated by two metrical positions resulting in a ternary rather than binary subdivision.

3.0.1 "The Farmer in the Dell"

		x			x			x			x
	x	x	x	x	x	x	x	x	x	x	x
	C	F		F	F		F	F			
	The	far-		mer	in		the	dell.			

As is the case with "The Drunken Sailor," the tune of subsequent verses may depart from the original:

3.0.2

		x			x			x			x
	x	x	x	x	x	x	x	x	x	x	x
	C	F		ø	F		F	F			
	The	rat			eats		the	cheese.			

Of the attested verses, these departures tend to be less significant than in "The Drunken Sailor." Most follow 3.0.2 in deleting the third syllable of the original text resulting in an unoccupied position indicated by ø. It is, however, fairly easy to compose acceptable verses which augment the original count with several additional syllables:

3.0.3 Acceptable constructs

		x			x			x			x
	x	x	x	x	x	x	x	x	x	x	x
	C	F	F	F	F	F	F	F			
a)	E	liz	a	beth	eats		the	cheese.			
b)	The	far		mer	dan	ces	a	jig.			
c)	E	liz	a	beth	dan	ces	a	jig.			

Variants which insert an event prior to the first event of the form are also, it appears, acceptable. Replacing “Elizabeth” by “Anastasia” in 3.0.2 results in such a setting.

3.0.4¹¹

			x			x			x			x
x	x	x	x	x	x	x	x	x	x	x	x	x
C	C	F	F	F	F	F	F	F	F	F	F	F
An	a	stas	i	a	eats	the	cheese.					

These are counterposed against constructs which are clearly unacceptable

3.0.5

			x					x					x
	x	x		x	x	x	x	x	x	x	x	x	x
	C	F		F	F	F	F	F	F	F	F	F	F
a)*	Em	il		y	takes	the	child.						
b)*	John	-			takes	the	child.						
c)*	∅	John			takes	the	child. ¹²						

It is worth re-emphasizing here that the objectionable features in 3.0.5 are of two fundamentally different sorts. 3.0.5 a) contains stressed-unstressed first and second syllables assigned to weak-strong metrical locations. It is therefore an unproblematic instance of an unacceptable interaction of text and tune, namely a stress-meter mismatch as discussed in 1.1. 3.0.5 b) manifests a different type of forbidden interaction: not a stress mismatch but, as discussed in 1.3, an improperly located melisma—namely the initiation of a melisma on a relatively weak metrical position.

¹¹ Again, the grid is reoriented so as to make available the additional event at the beginning of the form.

¹² Some readers may not find immediately apparent that the deletion of the initial event will produce an unacceptable variant. That the variant is, in fact, unacceptable can be seen by substituting a text such as “John was in the dell” which requires deleting the initial anacrusis not only from the initial phrase of the song but from the subsequent parallel phrase corresponding with the repetition of the text:

	x			x				x				x
	x	x		x	x	x	x	x	x	x	x	x
x x	x x x	x x x	x x x	x x	x x x x	x x x	x x x x					
(C) F	F F F	F		(G) A	A A A	A						
The far	mer in the	dell,		the far	mer in the	dell						
	John was	in the	dell,		John was	in the	dell.					

As was the case with 2.2 b above, 3.0.5 c) represents a deviant text setting of a fundamentally different sort from those identified in 3.0.5 a) and b). To reiterate the discussion in 2.2 above, the offending feature does not reside in an unacceptable interaction between text and tune, but rather is inherent in the tune itself. Whereas 3.0.5 a) and b) make use of one of the acceptable variants of the tune,¹³ no reasonable setting can allow the truncation of the first event and be considered an acceptable variant of the original. 3.0.5 c) is therefore entirely reasonable as a verse of a song, but not as a verse of the particular song we have been considering.

3.0.5 c) is only one of several distributions of events on the twelve positions of the metrical grid which would not be heard as acceptable variants of the tune, regardless of the text assigned to them. Others include:

3.0.6

		x			x			x			x	
	x	x	x	x	x	x	x	x	x	x	x	x
a)	C	F	F	F		F	F					
b)	C	F		F	F		F	F				
c)		F		F	F		F	F		F	F	

3.0.6 a) is identical to the original save for the shifting of the third event one position to the left. This simple alteration results in syncopations which clearly render the tune unacceptable as a variant strophe. The same goes for 3.0.6 b) which shifts the final two events rightward. Finally, 3.0.6 c) is unacceptable for two reasons: as already noted, it deletes the initial anacrusis from the original and it appends additional events to the end of the original. But while 3.0.6 c) is an unacceptable variant of "The Farmer in the

¹³These variants are actually attested in the song: 3.0.5a) occurs in our example 3.0.1 above, and 3.0.5b) occurs in 3.0.2.

Dell", that does not mean it is an unacceptable tune in its own right. Indeed it is the first phrase of a very well known American children's song.¹⁴

3.1. Similarity Metric-Formal Characteristics of Representation

At this point it is worth recapitulating our observations with respect to the characteristics of positions within the metrical grid which define acceptable variants: grid positions are either a) vacant or occupied and this occupancy or vacancy status is either b) optional or mandatory. In what follows, we will identify the four associative categories which this organization gives rise to by the following numerical code:

3.1.0

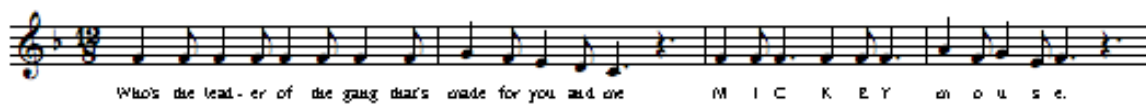
- 1 Mandatory occupancy : occupied in original and in all acceptable variants
- Optional vacancy: occupied in the original; may be deleted in acceptable variants.
- + Optional occupancy: vacant in the original; may be occupied in acceptable variants.
- ∅ Mandatory vacancy: vacant in original and in all acceptable variants.

What we will call the similarity metric associates each metrical position with one of these categories. Applying this taxonomy to "The Farmer in the Dell" yields the following organization:

3.1.1

		x			x			x			x
x	x	x	x	x	x	x	x	x	x	x	x
C	C	F	F	F	F	F	F	F	F	∅	∅
+	1	1	+	-	1	+	-	1	∅	∅	∅
	The	far		mer	in		the	dell.			

¹⁴ "The Mickey Mouse Club Song" (1955)



It will be seen that the pattern of mandatory occupancy and vacancy in 3.1.1 excludes as possible variants a large fraction of the 2^{12} mathematically possible arrangements of notes within the metrical grid. Those which are rejected, whether they are in themselves natural or irregular, are predicted to be heard as insufficiently similar to the original to be construed as strophic variants.

3.2 Similarity Metric: Derivation

The similarity metric in 3.1.1 represents, ultimately, a claim with respect to what constitutes a particular form of musical knowledge, specifically, what it means to "know" a tune and the form in which this knowledge consists. And in this respect the claim which it makes is somewhat counter-intuitive. It is generally assumed that knowing a melody means knowing the notes. Or, more technically, having access to a mental representation consisting of a string of pitch events assigned to positions on the metrical grid. The similarity metric suggests is that this commonsensical view is overly simplistic. Rather, knowledge of a tune encompasses not only a particular sequence but an awareness of the class of forms which this sequence can take. Thus, as discussed above, knowledge of "The Drunken Sailor" does not simply consist of knowledge of 2.1 a) (the original form) but the capacity of the tune to take the forms specified in 2.1 b) and c) among others. Or, put slightly differently, knowing a tune means in addition to having access to the original form, an awareness of the distribution of optional and mandatory positions which define acceptable variants as these are specified in the similarity metric.

Whatever form this knowledge takes, it clearly does not entail rote learning of the characteristics associated with each position of each variant of each tune. Even a listener

who has never heard 3.0.2 is able to recognize it as an acceptable variant of "The Farmer in the Dell" compared to the unacceptable constructs in 3.0.6. S/he is also likely to have fairly clear intuitions that the constructs in 3.0.3 are possible. Thus, it must be the case that class of acceptable variants are not learned but are inferred from the structural characteristics of the original.

Various aspects of musical intuition are implicated in the listener's competence along these lines. Our objective in the following is to make this competence explicit. Some insight on one aspect of this competence is provided by the fact that of the settings we have considered so far, neither of the optional categories - or + occurs on metrically strong positions. Or, put in more intuitively recognizable terms, events which can embellish the structure of existing tunes or which can be deleted without inducing unacceptable distortions tend to occur on weak positions. Strong positions, on the other hand, are accorded psychological prominence in the minds of listeners with the result that an original event occurring on a strong position cannot be deleted in a subsequent variant. This leads us to state the first of a sequence of rules which will determine the form of the similarity metric.

Similarity Metric Assignment Rule (hereafter SMAR) 1 (meter): Given the original form of a strophic song, assign category 1 to all strong metrical positions occupied in the original.

The term "strong" here is somewhat ambiguous. We will assume, for the moment, the definition adopted by Dell and Halle (2005) of strong positions as "tactus" level positions, meaning those locations which a listener would be likely to snap his or her

fingers in relation to, or according to which dance steps might be choreographed.¹⁵ In some cases, this choice is unambiguous. Thus, the tactus positions of "The Farmer in the Dell" are those belonging to the level indicated in 3.2.1. Those above and below this level are not viable candidates for the location of the tactus.

3.2.1

```

      x           x           x           x
      x     x     x     x     x     x     x     x (tactus)
x x x x x x x x x x x x x x x x x x x x x x x
  C F   F F   F F           G A   A A   A A

```

Also unproblematic is "The Drunken Sailor" which requires two levels below the tactus to account for the locations of all events:

3.2.2

```

      x           x           x           x           x (tactus)
      x           x           x           x           x
      x     x     x     x     x     x     x     x     x     x     x     x
      A           A     A     A           A     A     A           D     F     A
What           shall we do           with a drunk           en           sail           or

```

In some cases, the location of the tactus is somewhat provisional. Certain tunes can be sung at a relatively wide variety of tempi and this choice can force a reorientation of the tactus from a higher to lower metrical level. As we shall see, the location of the tactus can have consequences in determining what metrical positions are available for occupancy by variant forms.

Applications of SMAR 1 result in the following positions of "The Drunken Sailor" being designated as mandatory:

3.2.3

```

      x           x           x           x           x (tactus)
      x           x           x           x           x
      x     x     x     x     x     x     x     x     x     x     x     x
      A           A     A     A           A     A     A           D     F     A
SMAR 1   1           1           1           1           1

```

¹⁵ For a good definition and brief discussion of the term see London (2005).

A second constraint on the form taken by the similarity metric appears to derive from the pitch structure of the original melody. As we observed in 2.2 b, none of the notes of the final arpeggiated d minor triad at the end of the first phrase of "The Drunken Sailor" can be eliminated from acceptable variants. As a provisional explanation, we will assume here that the changes in pitch in events occurring on these positions is what is responsible for these positions being assigned mandatory status.

The song "au clair de la lune" provides further evidence for the influence of pitch on experienced listeners' judgments of acceptable and unacceptable variants of an original tune as these are represented in the similarity metric. The asterisks in 3.2.4 indicate events implicated in a pitch change, which we will define here as any pair of adjacent events assigned to different pitches.

3.2.4 Au clair de la lune

x		x		x		x		x		x		x		x		x		x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
F	F	F	G	A		G		F	A	G	G	F						
		*	*	*		*		*	*	*	*	*						
1	-	1	1	1	+	1	+	1	1	1	-	1	∅	∅	∅	∅ ¹⁶		

With one exception which we will return to later, the next to last event assigned to pitch G, it will be seen that all positions associated with changes in pitch are mandatory-that is, they are required to remain occupied in all variants and are therefore assigned to category 1 in the similarity metric. This observation suggests to the following assignment rule:

SMAR 2 (pitch change-first version): Assign category 1 to both members of a pair of adjacent events having different pitches.

¹⁶ We observe here that the categories assigned to metrical positions indicate the observed facts-namely, the intuitions of experienced listeners with respect to mandatory occupancy and vacancy. Our objective, which we are in the midst of pursuing, is to provide a formal account of these intuitions.

The two rules applied in sequence generate the following partial similarity metric applied to the metrical locations of "au clair de la lune." First, SMAR 1 assigns mandatory positions to strong beats as follows:

3.2.5 Au clair de la lune

		x		x		x		x		x		x		x		x	(tactus)	
		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
		F	F	F	G	A		G		F	A	G	G	F				
SMAR 1	1			1		1		1		1		1		1		1		

SMAR 2 then applies so as to assign mandatory status to all metrical positions involved in changes of pitch:

3.2.6 Au clair de la lune

		x		x		x		x		x		x		x		x	(tactus)	
		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
		F	F	F	G	A		G		F	A	G	G	F				
SMAR 1	1			1		1		1		1		1		1				
SMAR 2				1	1	1		1			1	1	1	1				

With the one previously noted exception, SMARs 1 and 2 correctly assign the appropriate mandatory positions. The note in question is the G preceding the final F. This is not, in fact, mandatory since it can be deleted in acceptable settings: A text such as "John walked in the moonlight when he came home." is assigned a setting which deletes the event assigned to the position in question in the original. This is indicated by parentheses in the following:

3.2.7

		x		x		x		x		x		x		x		x	
		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
		F	F	F	G	A		G		F	A	G	(G)	F			
a.		John	walked	in	the	moon	light	when	he	came	home						
b.		Au	clair	de	la	lu	ne	mon	a	mi	pier	rot					

Since 3.2.7 a is an entirely acceptable variant of the original indicated in 3.2.7 b, it is clear that the event in question can in fact be deleted from the original. It must be the case then that the change in pitch which would make the event mandatory according to the current version of SMAR 2 is not relevant here.

Two factors appear to be involved in correcting this result. First, while in the original tune there is a pitch change between the next to last event G and the event on its right (the final F), the G shares the same pitch with the event on its left (that is, it is preceded by another G). Secondly, the optionally deleted G is assigned to a weak metrical position in the original. It is assigned only a single x above it on the metrical grid while those its left and right are associated with a column having two x's.

These observations lead to the conclusion that pitch change, SMAR 2, is required to apply only when two conditions are met: 1) when the event in question is the weak member of a strong weak pair of events and 2) when both events of the strong weak pair are assigned to different pitches. Since the second condition is not met by the next to last event of "au clair de la lune", it is not mandatory within the similarity metric. These observations lead to the following revision of SMAR 2.

SMAR 2 (pitch change-revised): If X and Y are adjacent events assigned to different pitches, X precedes Y, and X is stronger than Y, assign category 1 to Y.

Informally stated, the SMAR 2 scans for adjacent strong-weak pairs having distinct pitches and assigns category 1 to the right member of this pair. With this revision SMARs 1 and 2 apply to produce the following partial similarity metric:

3.2.8

Application of SMAR 1 and 2:

		x							x										x	
		x		x		x		x		x		x		x		x			x	(tactus)
		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
		F	F	F	G	A		G		F	A	G	G	F						
SMAR 1		1		1		1		1		1		1		1						
SMAR 2					1			1							1					

Since SMARs 1 and 2 impose only a single category, the mandatory occupancy category 1, the similarity metric they assign is incomplete. We need additional rules to assign the remaining categories, -, + and \emptyset . For the assignment of the latter, we propose a third assignment rule which is suggested by the fact that the location of mandatory vacancies appears to be related to a relatively extended absence of events or a temporal gap which is heard as marking the right edge of a melodic group. The gap in temporal proximity divides the first twelve notes of the "The Farmer in the Dell " into what are clearly heard as two groups of six events as indicated by the vertical lines (|) in the following.

3.2.9¹⁷

		x		x		x		x		x		x		x		x		x		x							
		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x						
			C	F		F	F		F	F	\emptyset	\emptyset	\emptyset	\emptyset		G	A		A	A		A	A	\emptyset	\emptyset	\emptyset	

Any setting which excessively encroaches on the sequence of unoccupied positions will be heard as violating the integrity of the tune and will be rejected as a variant. One of these was the setting 3.0.6 mentioned above which assigns an event to the first of the four unoccupied positions:

¹⁷ The exact placement of the grouping boundary marks in relation to the grid need not concern us for the moment. We will return to this question presently.

3.2.10*¹⁸

```

      x      x      x      x      x      x      x      x
x x x x x x x x x x x x x x x x x x x x x x x x
  C F   F F   F F F ∅ ∅ ∅ G A   A A   A A A ∅ ∅

```

However, as we saw in 3.0.4, the final position of the gap can be occupied when it functions as an anacrusis to the second group (provided that the parallel first phrase has also been initiated with an anacrusis):

3.2.11¹⁹

```

      x      x      x      x      x      x      x      x
x x x x x x x x x x x x x x x x x x x x x x x x
  C C F   F F   F F ∅ ∅ ∅ G G A   A A   A A ∅ ∅ ∅
  |                   |
anacrusis             anacrusis

```

These examples indicate that the sequence of empty positions defining the right edge of a group extends up to and including the final strong position. The subsequent weak position is understood as belonging to the subsequent group, thus the grouping boundary should be located at this point-i.e. after the strong position and before the subsequent weak position:

3.2.12

```

      x      x      x      x      x      x      x      x
x x x x x x x x x x x x x x x x x x x x x x x x
|∅ C F   F F   F F ∅ ∅ ∅|∅ G A   A A   A A ∅ ∅ ∅|

```

To maintain the integrity of the group, the vacant positions at the right edge of the group must remain unoccupied, that is, they must be designated as mandatory vacancies, category ∅. The initial position of each group as we noted, is optionally occupied and must be designated category +.

¹⁸ The unacceptability of this sequence can be confirmed by substituting a text such as "The farmer ate the carrot" which imposes a stress-unstressed couplet-i.e. a feminine ending-at the end of the group. This variant is, interestingly, taken as acceptable within prosodic form in poetic traditions. (See, for example, Malof, 1970) We will return to this question in a subsequent chapter.

¹⁹ Those having difficulty deciphering the notation in 3.2.11 may find useful substituting the text "Arizona won the game" into "The Farmer in the Dell". This text will require the indicated musical sequence.

It will be seen in the following that these four rules generate a similarity metric which reasonably closely models the competence of most minimally competent musicians in designating acceptable variant forms.

3.3 Similarity Metric: Applications

To demonstrate the use of the SMAR system, we return first to the initial phrase of "The Farmer in the Dell."

3.3.1 "The Farmer in the Dell" (original tune)

		x			x			x			x
x	x	x	x	x	x	x	x	x	x	x	x
	C	F		F	F		F	F			

At the outset, SMAR 1 assigns category 1 to all occupied strong positions.

3.3.2 SMAR 1

			x			x			x			x
	x	x	x	x	x	x	x	x	x	x	x	x
		C	F		F	F		F	F			
SMAR 1			1			1			1			

SMAR 2 requires two events assigned to different pitches where the first event is stronger than the second. Since this condition is not met in "The Farmer in the Dell," the rule does not apply.

3.3.3 SMAR 2 (no applications)

			x			x			x			x
	x	x	x	x	x	x	x	x	x	x	x	x
		C	F		F	F		F	F			
SMAR 1			1			1			1			
SMAR 2												

As previously discussed, SMAR 3 assigns mandatory status to right and left edges of groups whose boundaries are indicated by |:

3.3.4 SMAR 3

			x			x			x			x
	x	x	x	x	x	x	x	x	x	x	x	x
		C	F		F	F		F	F			
SMAR 1			1			1			1			
SMAR 2												
SMAR 3		1								∅	∅	∅

Finally, SMAR 4 makes the default assignment of optional categories to the remaining positions:

3.3.5 SMAR 4

			x			x			x			x
	x	x	x	x	x	x	x	x	x	x	x	x
	C ²⁰	C	F		F	F		F	F			
SMAR 1			1			1			1			
SMAR 2												
SMAR 3		1								∅	∅	∅
SMAR 4	+			+	-		+	-				
output:	+	1	1	+	-	1	+	-	1	∅	∅	∅

This, as discussed previously, predicts the correct forms-i.e. the range of forms which are associated with strophic variants of the original.

The following examples apply the SMAR system to the melodies we have been concerned with so far.

²⁰ The assignment of pitch to grid positions which have been designated by SMAR 4 as optionally occupied (+) will be discussed in section 3.5.

3.3.6 "au clair de la lune"

```

x           x           x           x
x         x   x       x   x   x   x   x   x   x
x  x   x  x  x  x  x  x  x  x  x  x  x  x  x
F  F   F  G  A           G   F  A  G  G  F
Au clair de la lu     ne     mon a mi pierrot21

```

SMAR applications:

```

           x           x           x           x
           x         x   x       x   x   x   x   x   x
           x  x  x  x  x  x  x  x  x  x  x  x  x  x
SMAR 1 | F  F  F  G  A           G |22 F  A  G  G  F           |
SMAR 2           1           1           1
SMAR 3 1
SMAR 4 - + + -

```

combined SMAR output:

```

x           x           x           x
x         x   x       x   x   x   x   x   x
x  x   x  x  x  x  x  x  x  x  x  x  x  x
Au clair de la lu     ne     mon a mi Pierrot
F  F   F  G  A           G   F  A  G  G  F
1  -   1  1  1  +  1  +  1  1  1  -  1  0  0  0

```

3.3.7 "The Drunken Sailor"

```

           x           x           x           x
           x         x   x       x   x   x   x   x   x
x  x   x  x  x  x  x  x  x  x  x  x  x  x  x
A         A  A  A           A  A  A           D           F  A
What      shall we do with a drunk en sail or

```

SMAR applications:

```

           x           x           x           x
           x         x   x       x   x   x   x   x   x
x  x   x  x  x  x  x  x  x  x  x  x  x  x
A         A  A  A           A  A  A           D           F  A
SMAR 1           1           1           1           1
SMAR 2           1           1           1
SMAR 3
SMAR 4 + + - - + - - + + +

```

²¹ We include the text here as a mnemonic to those who are not able to recognize the structure of the tunes from the musical form alone. We reiterate that the associated text has no bearing on the independent musical structure which the similarity metric is formulated to explain.

²² The grouping boundary at this location will not be considered as salient for SMAR 3 for reasons discussed in section 3.7.

combined SMAR output:

```

      x           x           x           x
      x           x           x           x
x   x   x   x   x   x   x   x   x   x   x   x   x   x   x
    What       shall we do       with a drun       ken       sai       lor?
      A         A   A   A         A   A   A         D         F         A
+   1   +   -   -   1   +   -   -   1   +   -   +   1   +   -

```

3.3.8 "Frère Jacques"

```

      x           x           x           x
      x           x           x           x
x   x   x   x   x   x   x   x   x   x   x   x   x
    Frè     re   Jac   ques   frè     re   Jac   ques
      C         D         E         C         C         D         E         C

```

SMAR applications:

```

      x           x           x           x
      x           x           x           x
      x   x   x   x   x   x   x   x   x   x   x   x   x
|   C         D         E         C         C         D         E         C   |23
SMAR 1   1           1           1           1           1           1           1
SMAR 2           1           1           1           1           1           1
SMAR 3           1           1           1           1           1           1
SMAR 4 +   +   +   +   +   +   +   +   +   +   +

```

SMAR output:

```

      x           x           x           x
      x           x           x           x
x   x   x   x   x   x   x   x   x   x   x   x   x
    Frè     re   Jac   ques   frè     re   Jac   ques
      C         D         E         C         C         D         E         C
+   1   +   1   +   1   +   1   +   1   +   1   +   1   +

```

3.4 Discussion of Results

An arithmetic consequence of having a class of variants defined by one or more optional positions is that the number of variants defined by the similarity metric will increase exponentially according to the number of optional metrical positions. Thus, the four optional positions of "au clair de la lune" allow for sixteen possible melodies resulting from the four optional positions indicated by categories - and +:

²³ Neither grouping boundary will be considered salient for the assignment of SMAR 3. See section 3.7 for discussion.

3.4.1²⁴

	x			x				x				x					
	x	x		x	x		x	x	x	x	x	x	x	x		x	
	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	1	-	1	1	1	+	1	+	1	1	1	-	1	∅	∅	∅	
a)	F	F	F	G	A	A	G		F	A	G	G	F				
b)	F		F	G	A	A	G		F	A	G	G	F				
c)	F	F	F	G	A		G		F	A	G	G	F	(original)			
d)	F	F	F	G	A	A	G		F	A	G		F				
e)	F	F	F	G	A		G		F	A	G		F				
f)	F		F	G	A	A	G		F	A	G		F				
g)	F		F	G	A		G		F	A	G	G	F				
h)	F		F	G	A		G		F	A	G		F				
i)	F	F	F	G	A	A	G	G	F	A	G	G	F				
j)	F		F	G	A	A	G	G	F	A	G	G	F				
k)	F	F	F	G	A		G	G	F	A	G	G	F				
l)	F	F	F	G	A	A	G	G	F	A	G		F				
m)	F	F	F	G	A		G	G	F	A	G		F				
n)	F		F	G	A	A	G	G	F	A	G		F				
o)	F		F	G	A		G	G	F	A	G	G	F				
p)	F		F	G	A		G	G	F	A	G		F				

The forms in 3.4.1 are similar to the original in the sense that all will sound reasonably close to the original. This will be seen to be the case when these are compared to forms which the similarity metric does not generate. In the following example, unoccupied mandatory locations, those which are required to be occupied by events but which are not in the setting are indicated by * while occupied positions are required to remain vacant are indicated by the offending event appearing in italics:

3.4.2*

	x			x				x				x					
	x	x		x	x		x	x	x	x	x	x	x	x	x		x
	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	1	-	1	1	1	+	1	+	1	1	1	-	1	∅	∅	∅	
a)	F	F	F	G	A	A	*		F	*	*	G	F				
b)	*	F	*	G	A	A	G		F	A	*	G	*				
c)	F	F	F	G		A	G		F	A	G	G	F	<i>F</i>			
d)	F		F	*	A	*	G		F	A	G	G	F				

While the subjective impression of the similarity or lack thereof to the original is significant, the relevant evaluation of these forms can be made on more specific grounds.

²⁴ The grids here do not indicate the pitches to be assigned to the optional events of some of the variants. We will deal with pitch assignment in section 3.5.

All of the forms in 3.4.1 will appear as variants of "au clair de la lune" when texts require them. Thus, for example, the text "John was tired so he went to bed" will require 3.4.1 g.

3.4.3²⁵

```

      x           x           x           x
      x     x     x     x     x     x     x     x     x
      x x x x x x x x x x x x x x x x
      F     F G A     G     F A G     G F
John   was - ti     red   so he went to bed.

```

On the other hand, all of the forms in 3.4.2 are definitively excluded as variants of "au clair de la lune." That is, they will not appear as variants even when the text selected allows for an acceptable interaction of tune and text. Thus, for example, the tune 3.4.2 d will make for an appropriate match for the previous text but it is not an acceptable variant of the tune:

3.4.4*

```

      x           x           x           x
      x     x     x     x     x     x     x     x     x
      x x x x x x x x x x x x x x x x
      F     F ø A     G     F A G     G F
John   was  ti     red   so he went to bed.

```

The explanation for the rejection of 3.4.4 is specified in the similarity metric: the G in the fourth metrical position must appear in all variants, either assigned to a syllable or when no syllable is available, as is the case here, as a melisma. The vacancy, while not adversely affecting the relationship of text and tune, results in a tune which cannot be construed as a variant.

²⁵ We recognize an additional acceptable setting associating the text and tune is available:

```

      x           x           x           x
      x     x     x     x     x     x     x     x     x
      x x x x x x x x x x x x x x x x
      F     (F) G A     G     F A G     G F
John   was  tir  ed   so he went to bed.

```

This form will be discussed in 3.6 below.

To reiterate a point made at the outset, the similarity metric defines the class of available independent musical forms. It has nothing to say about interactive form, namely that inhering in the relationship between text and tune, nor does it have anything to say about *the text* as an independent form. Of course, it does not follow that all grammatical utterances will work equally well as texts associated with a tune. Clearly, a text having roughly the same number of syllables as notes in the original will, all things being equal, be preferable to one which has a great deal more or fewer. Moreover, even texts having the same number of syllables can allow for a range of highly natural to unnatural settings depending on the distribution of stresses and prosodic boundaries. For example, while the texts "John did not pay alimony recently" and "John was in a frenzy when he came back home" both have eleven syllables only the latter is easily set to the tune of "au clair de la lune." The former appears to resist any viable setting even given the large number of possible variants. Of those shown in 3.4.6, 3.4.6 a is fully acceptable, whereas the remaining arrangements are at best awkward (3.4.6 b-d) and at worst completely impossible (3.4.6 e).

3.4.6

x				x					x				x		
x		x		x	x			x	x			x	x		x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
1	-	1	1	1	+	1	+	1	1	1	-	1	∅	∅	∅
F	F	F	G	A	(A)	G		(G)	F	A	G	G	F		
a)	John	was	in	a	fren	zy		when	he	came	back	home.			
b)*	John	did	not	pay	[al	i mon	y]	re	-	cent	ly.				
c)*	John	did	not	pay	[al	i	i	mon	y]	re	cent	ly.			
d)*	John	did	not	pay	[al	i		mon	y]	re	cent	ly.			
e)*	John	did	not	pay	[al	i mon		y]	re	cent	ly.				

The underlying explanation for these intuitive judgements is that each contains a violation of one or more of the principles governing acceptable text to tune interactions discussed previously. These include (as indicated) constituency mismatches resulting from the word "alimony" being separated between two groups, stress mismatches, or

both. It is due to these violations resulting from any possible arrangement of the text that the text itself seems to be inadmissible.

This brings up a larger point which is that when one speaks of "unmetrical" texts in the context of text setting, one is referring to examples such as 3.4.6 b-e, i.e. texts which cannot be accommodated within a particular strophic song form. However, it is important to recognize that the basis of the rejection of these settings is interactive-it is not a characteristic of the text per se, but the text in relation to the tune. In other words, it is an a posteriori constraint on the interaction, not an a priori constraint on the text. The closely related notion of unmetrical texts for so called "metrically rigid"²⁶ or "strict-stress"²⁷ verse patterns should, we believe, be seen in an analogous light in many instances. Metricality within these verse types is not an a priori characteristic of texts but an a posteriori characteristic inhering in the arrangement of texts and the underlying pitchless tune which defines a given poetic meter. We will return to this subject in subsequent research.

3.5 Pitch Assignment

No mention has been made so far as to how pitch is assigned to variant forms defined by the similarity metric. This is because, for the most part, no discussion is necessary: the pitch assigned to grid locations follows from the definition of strophic form which dictates that initially occupied positions, those assigned either category 1 or -, maintain the pitch they were initially assigned while vacant locations assigned category \emptyset do not require pitch. A specification of pitch is required only for optionally occupied locations, those assigned category +, since they are vacant in the original and potentially occupied

²⁶ Oehrle (1989)

²⁷ Tarlinskaya (1993)

in variants. It will be seen that, of the tunes we have considered so far, the procedure is easily stated: optional locations inherit pitch from the nearest assigned location to the left. Thus, as indicated in 3.5.1 the pitches assigned to the two category + positions of "au clair de lune" are the same as those assigned to the preceding locations. That any other pitch at this location will sound dramatically wrong or at least somewhat awkward can be verified by substituting alternate pitches into the parenthesized locations.

3.5.1 Leftward pitch inheritance

x		x		x		x		x		x		x		x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
1	-	1	1	1	+	1	+	1	1	1	-	1	∅	∅
F	F	F	G	A<(A)	G<(G)	F	A	G	G	F				

A difficulty with leftward inheritance will be seen to arise when the optional event to be assigned pitch is the initial event of a sequence. For example, in the "Farmer in the Dell," while the non-initial pitches of category + positions are correctly accounted for by leftward inheritance, the initial location has no left position to inherit pitch from:

3.5.2

		x			x			x			x
x	x	x	x	x	x	x	x	x	x	x	x
+	1	1	+	-	1	+	-	1	∅	∅	∅
<(?)	C	F	<(F)	F	F	<(F)	F	F			

An ad hoc solution would specify rightward inheritance for initial grid positions. While this solution would correctly predict the pitch associated with the first position of "The Farmer in the Dell," it fails to correctly predict the optional anacrusis to "Frère Jacques" which Morgan and Janda's informants appended to the original:

3.5.3

	x				x				x				x		
	x		x		x		x		x		x		x		x
	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
(G)>	C		D		E		C		C		D		E		C
	Frè		re		Jac		ques		frè		re		Jac		ques
	My bro		thers		dog		-		keeps		on		bark		ing

Here, according to the informants, the anacrusis is assigned to the fifth scale degree-the G below the tonic downbeat C which initiated the original tune.

The perfect fourth upward leap appears to be the normative configuration for optional anacrusis preceding scale degree 1. Other downbeat pitches require different pitch configurations. For example, the initial downbeat assigned to scale degree 5 in "The Drunken Sailor" requires an anacrusis assigned to the same pitch:

3.5.4

			x				x				x				x
		x	x		x		x		x		x		x		x
	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
(A)	A	A	A	A	A	A	A	A	A	A	D	F	A	A	A
a.	What	shall	we	do	with	a	drunk	en	sail	or					
b.	Re-lease	him	there	and	make	him	bail	her							

Finally, there are circumstances in which the pitch does not appear to be easily predictable, most notably when the tune begins on scale degree 3 (E in the following C major tune):

3.5.5

			x				x				x
	x	x	x	x	x	x	x	x	x	x	x
	+	+	1	+	+	1	+	+	1	∅	∅
		?	E			D			C		
a.		Three		blind		mice.					
b.	El	laine		walked		home.					

Unlike the preceding cases, there does not appear to be a clear preference with respect to what pitch should be assigned to the anacrusis of 3.5.5: scale degrees 1 (C), 3 (E) and, possibly, 5 (G) all seem more or less acceptable with the proviso that they are not higher

than the main pitch: that is, if the first syllable of "Elaine" in 3.5.5 b were assigned to the C, E or G above the E on the strong beat. The operative intuition appears to be that any pitch which would result in a leap to the downbeat from above sounds unnatural as a possible anacrusis and is rejected on this basis. Furthermore, while we will not examine these specifically, it will be seen that there are relatively strong intuitive preferences militating against most non-scalar tones serving as an anacrusis, (e.g. C#, Gb etc.) as well as scale degrees other than those just mentioned, (e.g. 2, 6 and 7).

Rather than positive guidance, the ear seems to be responding to a variety of interacting negative intuitive preferences in judging what can constitute the correct pitch assigned to the anacrusis. A procedure which successfully balances these intuitions in assigning pitch is therefore likely to require a relatively complex calculation beyond the scope of the present discussion. For our purposes, we will content ourselves with the informal procedure for pitch assignment outlined above, namely leftward inheritance for all non-initial events on the similarity metric. The pitch assigned to anacruses will be determined on a case-by-case basis guided by the rough outlines just mentioned.

3.6 Lacunae: Fusion

Before closing out our discussion of the similarity metric, it should be understood that what is presented above is a first approximation. It is not fully satisfactory in that one can find instances where it both under and overgenerates: forms which seem intuitively reasonable are ruled out and relatively unnatural forms are predicted to be acceptable by the model.

the attested setting 3.6.1 which leaves this position vacant.

However, this gambit comes with a price in that it makes available three problematic settings, among them the two syncopated variants 3.6.4 a) and b). Still less acceptable is 3.6.4 c) which undermines the final stepwise descent to the tonic.

3.6.4

	x			x				x				x				(tactus)	
	x		x		x		x		x		x		x		x		x
	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	1	-	-	1	1	+	1	+	1	1	-	-	1	∅	∅	∅	
a)	F	F	F	G	A		G		F	A		G	F				
b)	F	F		G	A		G		F	A	G	G	F				
c)	F	F	F	G	A		G		F	A			F				

A possible solution suggested by Francois Dell (p.c) is to maintain the location of the tactus as in 3.6.2 but allow for leftward "fusion" of adjacent identical pitches. Fusion takes as its basis the intuition that subsequent repetition of pitches are often heard not as independent events but rather as continuations of the initial event. Thus, the F at the beginning of "au clair de la lune" may fuse together with rightward adjacent occurrences of the same pitch shown here in brackets:

3.6.5

x		x		x		x
x	x	x	x	x	x	x
[F	F	F]	G	A		G

Fusion of the bracketed unit will combine the three events into a single event initiated by the first occurrence of the pitch:

3.6.6

x		x		x		x
x	x	x	x	x	x	x
F		G	A			G

Fusion, as expressed within the similarity metric, requires that any mandatory status initially assigned to reiterations of repeated pitch events may be redesignated as optional so that these attacks can be deleted in the variant. Thus, what was a strong position assigned to the third occurrence of the F in the original:

3.6.7

x		x		x		x
x	x	x	x	x	x	x
1	-	1	1	1		1
[F	F	F]	G	A		G

will be redefined as optional category -.

3.6.8

x		x		x		x
x	x	x	x	x	x	x
1	-	-	1	1		1
[F	F	F]	G	A		G

This allows for the optional deletion of the two repetitions of the similar pitch and the prediction of the attested setting 3.6.2.

The conversion of formerly mandatory positions occupied by repeated pitches into optional positions may only apply in limited circumstances. For example, the repeated pitches in the song "Frère Jacques" are not allowed to fuse together. The fused variant 3.6.9 b is clearly not acceptable.

	x			x			x			x			x			x
	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x
	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	Frè		re		Jac		ques		frè		re		Jac		ques	
a	C		D		E		[C		C]		D		E		C	
b*	C		D		E		C		ø		D		E		C	

Similarly, fusion must be prevented from applying to the bracketed pitches in 3.6.10a to create the unacceptable variant 3.6.10 b.

3.6.10²⁸

		x			x			x			x				
		x		x	x		x	x		x		x		x	
	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
		What		shall	we	do		with	a	drun		ken		sai	lor?
a		A		A	[A	A]		A	[A	A]		D		F	A
b*		A		A	A	∅		A	A	∅		D		F	A

The deleted strong positions indicated in 3.6.9 and 3.6.10 b by ∅ resulting from fused similar pitches are clearly not allowed in an acceptable variant of either tune.

What is required is to identify the conditions under which fusion may or may not apply.

What appears to be relevant is grouping structure, specifically the division of the eight note sequence into two identical subgroups containing the pitches C, D, E and C:

3.6.11

		x			x			x			x				
		x		x	x		x	x		x		x		x	
	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
		Frè		re		Jac	ques		frè		re		Jac	ques	
a.		C		D		E	C		C		D		E	C	

While it is less immediately apparent "The Drunken Sailor" is also divided into subgroups defined by two identical sequences. These are defined, as indicated, by a long (L) and two short (S) events:

3.6.12

		x			x			x			x							
		x		x	x		x	x		x		x		x				
	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
		What		shall	we	do		with	a	drun		ken		sai	lor?			
		A		A	A		A		A	A		A		D		F	A	
		L		S	S	L		S	S									

²⁸ See Temperly (1999) for an interesting discussion of the availability of syncopated forms like these in the rock idiom. It is worth noting that an *unsyncopated* tune would be unavailable within this style as either as an original or as a variant based on the normative requirements of the style. This raises the interesting question as to what extent the variants selected by the similarity metric are stylistically contingent and which are universal similarity judgments independent of style. For the moment, we will leave this matter unaddressed.

A reasonable conclusion appears to be that identical pitches may not be fused when the two pitches are separated by a subgroup boundary as defined above. This observation suggest the following formulation of fusion as SMAR 5:

SMAR 5 (fusion): Given two adjacent identically pitched events X1 and X2 within the same subgroup of an original tune where X1 precedes X2, redesignate the location containing X2 as optionally vacant category 2.

SMAR 5 will result in allowing optional vacancy of the metrical position associated with fusion under the conditions defined by 3.6.2 but not under the conditions defined in 3.6.9 and 3.6.10. This admittedly provisional result at least succeeds in identifying some of the factors relevant in the intuitions involved in making these sorts of reductional²⁹ judgments.

3.7 Further Lacunae: Restrictions on Grouping Structure

Finally, we return to the incorrect prediction indicated in footnotes 17 and 18 made by SMAR 3 applied to "au clair de la lune" which is assigned the following grouping structure:

3.7.1

x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	F	F	F	G	A	G		F	A	G	G	F			
A	u	c	l	a	i	r	d	e	l	a	l	u	n	e	m

If SMAR 3 were to apply at the medial grouping boundary indicated, namely, to the immediate left of the seventh event, the result would be the assignment of a mandatory vacancy at the preceding weak position:

²⁹ See Lerdahl and Jackendoff (1983) chapter 5 and also Lerdahl (2000) for an introduction to the involved topic of musical reduction and its cognitive status.

3.7.2

```

      x      x      x      x      x      x      x      x
      x  x    x  x  x  x  x  x  x  x  x  x  x  x  x
      F  F    F  G  A      G      F  A  G  G  F
SMAR 3                                ∅

```

This would require that half of the settings which we claimed to be viable in 3.4.1 (3.4.1 i – p), are in fact unacceptable. This seems unlikely; for example, a text such as, "In a town near Washington he seemed to lose his way" seems to be fairly unproblematically assigned 3.4.1 i):

3.7.3

```

      x      x      x      x      x      x      x      x
      x  x  x    x    x  x  x  x  x      x  x  x    x  x  x  x
      F  F  F    G    A  A  G  G  F      A  G  G    F
      In a town near Washington he seemed to lose his way.

```

It appears that SMAR 3 needs to be blocked from applying at the location indicated in 3.7.2. This can be accomplished in two ways. One approach is to mandate a grouping analysis which situates the boundary to the immediate right of the sixth event of the original tune:

3.7.4

```

      x      x      x      x      x      x      x      x
      x  x    x  x  x  x  x  x  x  x  x  x  x  x
      | F  F    F  G  A      G |      F  A  G  G  F      |

```

This analysis is problematic, however, from the standpoint of constituency matching: If for example, the text "John went to the supermarket on his way to school" is inserted as shown in 3.7.5 the grouping boundary will intercede within the word "supermarket" and a constituency mismatch will result:

3.7.5

x x x x x x x x
 x x x x x x x x x x x x x x x x
 F F F G A A G | G F A G G F
 John went to the [supermarket] on his way to school.

A second possible solution is to limit the application of SMAR 3 exclusively to a particular variety of grouping boundary, namely that which results from temporal proximity (see Halle, forthcoming). Proximity groupings result from a particular event being relatively long compared to those adjacent to it. For example, the grouping structure assigned to "Hernando's Hideaway" referred to earlier will be seen to correspond with the longest events-those separated from the attack of the following event by at least four eighth notes.

3.7.6



Returning to "au clair de la lune," the temporal proximity condition does not impose the medial grouping boundary indicated in 3.7.4: the seventh event and the preceding sixth event in the original (assigned to the word "lune") are equally long in that both require two metrical locations. Neither is, therefore, relatively long compared to those adjacent to it.

On the other hand, the final event of the group-that assigned to the second syllable of "Pierrot"- requires four locations on the metrical grid. It is, therefore, as discussed in Halle (forthcoming) maximally long-i.e. longer than all others in the immediate vicinity.

And it is to the right of this location where the relevant grouping boundary needs to be located for the purpose of identifying mandatory vacancies in the similarity metric. It appears that SMAR 3 is only sensitive to this particular class of grouping boundary, that is, grouping resulting from proximity. SMAR 3 ignores group boundaries resulting from other musical factors most notably those deriving from rhythmic, pitch and timbral similarity.³⁰ This observation suggest the following revision of SMAR 3:

SMAR 3. (grouping-revised) Within each group, designate as mandatory a) the metrical position corresponding to initial event in the original and b) all metrical positions to the right of the final event of a group when this event is maximally long.

This revision allows for the proper assignment of a similarity metric to "au clair de la lune." It also addresses a similar problem noted in footnote 23 with respect to "Frère Jacques."

3.8 Conclusion

In concluding this discussion, it is worth pointing out a potential methodological/empirical problem with the class of data we have concerned ourselves with. The intuitive judgements of acceptability which we have relied on to make our determinations as to the structure of the similarity metric are not, in all cases at least, entirely firm and decisive. We have asserted that certain forms are variously acceptable or unacceptable, and while we have made occasional references to uncertainties in these judgements, it is clear that there is a substantial grey area of marginally unacceptable or acceptable settings which are difficult to differentiate from each other. Having said that, the above analysis is

³⁰ See Lerdahl and Jackendoff (1983) chapter 3 for discussion of these additional factors. The experimental work of Cohen and Frankland (2004) suggests that proximity is the primary factor in determining melodic grouping with other parameters exerting only a marginal influence.

primarily reliant on relatively clear cases. Generally, the problematic cases are at the periphery of the core facts that it attempts to explain.

Part of the empirical and technical difficulties just mentioned go with the territory, so to speak: judgments of similarity in cognitive domains while subjectively experienced as quick, easy and unproblematic are objectively far from trivial. The classic work of Carey (1985) on facial recognition is perhaps the best known demonstration that even comparatively simple similarity judgments are, from a computational standpoint, highly complex. There is good reason to expect, based on the preceding discussion, that even the relatively circumscribed similarity judgments implicated in the creation of strophic forms will require theoretical machinery of considerably greater complexity than that proposed above. That it gets us as far as it seems to, however, is reason to believe that additional progress can be made along the general lines proposed here.

In any case, our purpose here is to demonstrate the place of a similarity metric within the context of text setting and metrical form more generally. It seems clear that a place for it, in some form, is required. Following chapters will demonstrate the relevance of the similarity metric to a variety of metrical idioms.

References

Boone, Graeme M. (1999) *Patterns in play : a model for text setting in the early French songs of Guillaume Dufay*. Lincoln : University of Nebraska Press.

Carey, Susan (1985) *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Cohen, Annabel J and Bradley W Frankland, *Parsing of Melody: Quantification and Testing of the Local Grouping Rules of Lerdahl and Jackendoff's "A Generative Theory of Tonal Music"* in *Music Perception* 21:4 [Summer 2004] p. 499-543

Crocker, Richard (2005) Melisma in *The New Grove Dictionary of Music and Musicians*, on-line edition, Stanley Sadie and John Tyrrell eds. London: Oxford University Press

Dell, François, and John Halle (2005) Comparing musical textsetting in French and in English songs to appear in proceedings volume from the first international Paris conference on the typology of metrical forms.

Fehn, Anne C. and Rufus Hallmark, (1983) Text and music in Schubert's pentameter Lieder: a consideration of declamation' in *Studies in the history of music*, vol 1: music and language, New York: Norton

Gauldin, Robert (1995) *A Practical Approach to Sixteenth/Eighteenth-Century Counterpoint* Prospect Heights, Illinois: Waveland Press, Inc.,

Halle, John and Fred Lerdahl (1993) A generative textsetting model. *Current Musicology* 55:3-23.

Halle, John (1999) *A Grammar of Improvised Textsetting*. Ph.D. dissertation, Columbia University.

Halle, John (forthcoming) Constituency matching in metrical texts. To appear in proceedings of the University of Missouri Conference on Words and Music

Hayes, Bruce. (1984) The phonology of rhythm in English *Linguistic Inquiry* 15, 33-42.

Hayes, Bruce and Abigail Kaun (1996) The role of phonological phrasing in sung and chanted verse. *The Linguistic Review* 13:243-303.

Hayes, Bruce and Margaret MacEachern (1996) Are there lines in folk poetry?, *UCLA Working Papers in Phonology* 1:125-142.
[<http://www.linguistics.ucla.edu/people/hayes/metrics.htm>]

Hayes, Bruce and Margaret MacEachern (1998) Quatrain form in English folk verse. *Language* 74:473-507.

Kiparsky, Paul, (forthcoming) A Modular Metrics for Folk Verse. To appear in B. Elan Dresher and Nila Friedberg (eds.) *Formal approaches to poetry*. The Hague: Mouton,

Kiparsky, Paul and Gilbert Youmans, eds. (1989) *Rhythm and Meter*. Orlando, FL: Academic Press.

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

Lerdahl, Fred (2001) The Sounds of Poetry Viewed as Music *Annals of the New York Academy of Sciences* 930:337-354

Lerdahl, Fred (2001) *Tonal pitch Space* New York: Oxford University Press.

Liberman, Mark (1975) *The Intonational System of English*, Ph.D. dissertation, MIT. Published 1979: New York, Garland Publishing, Inc.

London, Justin (2005) Rhythm, Tempo and Tactus in The New Grove Dictionary of Music and Musicians, on line edition, edited by Stanley Sadie and John Tyrrell. London: Oxford University Press

Malof, Joseph (1970) Manual of English Meters. Bloomington: Indiana University Press.

Nespor, M. and Vogel, I. (1986) *Prosodic Phonology*, Foris: Dordrecht.

Morgan, Terrell A. and Richard D. Janda (1989) "Musically-conditioned stress shift in Spanish revisited: empirical verification and nonlinear analysis". In Carl Kirschner, and Janet Ann DeCesaris (eds.), *Studies in Romance Linguistics, Selected Proceedings from the XVII Linguistic Symposium on Romance Languages*. New Jersey: Rutgers University

Oehrle, Richard (1989). Temporal structures in verse design. In Kiparsky and Youmans (1989), pp. 87-119.

Palmer, C. & Kelly, M. (1992) Linguistic prosody and musical meter in song *Journal of Memory and Language*, 31, 525-542.

Selkirk, E.O. (1984) *Phonology and Syntax*, Cambridge, MA: MIT Press.

Stein, David, and David Gil (1980) Prosodic structures and prosodic markers. *Theoretical Linguistics* 7. 173-240

Stoquerus, Gaspar (1570: 1988) *De musica verbali libri duo* (Madrid: Biblioteca Nacional, Codex 6486); ed. and trans. Albert C. Rotola, *De musica verbali libri duo / Two Books on Verbal Music* Lincoln: University of Nebraska Press.

Tarlinskaja, Marina (1993) *Strict Stress-Meter in English Poetry Compared with German and Russian*. Calgary: University of Calgary Press.

Temperley, David. (1999) Syncopation in rock: a perceptual perspective. *Popular Music* 18:19-40.

Veltman, Joshua (2001) *Textual Accent And Mensural Rhythm In Sixteenth-Century Settings Of The "Pater Noster"* Master's thesis: The Ohio State University.