Elena Girardi & Ingo Plag

Metrical mapping in text-setting: Empirical analysis and grammatical implementation

Abstract

Linguists and musicologists share a persistent interest in text-setting, that is, in how the language of a poem is matched to musical structure (e.g. Halle & Lerdahl 1993, Kiparsky 2006, Hayes 2009). For example, it is generally assumed that in English vocal music, stressed syllables tend to fall on relatively strong beats of the musical meter (Palmer and Kelly 1992, Halle and Lerdahl 1993), but there is surprisingly little quantitative evidence for it (cf. Temperley & Temperley 2013: 521). In this paper we present a systematic quantitative investigation of how poetic meter (i.e. S and W positions of a metrical template) and syllabic prominence is matched to three dimensions of musical structure: musical meter, note value, and pitch. The corpus consists of three different *Lieder* by Haydn, based on three poems by Hunter, and three different settings of the same poem (*Annabel Lee* by Poe), with 161 lines overall. Using a set of constraints that evaluate the mapping of poetic/linguistic and musical structure, we are able to establish grammars that underlie particular text-settings. The songs in our sample vary in the probabilities with which particular constraints are violated. In all text-settings, constraints demanding the alignment of poetic-metrical positions with musical metrical strength are less easily violated than constraints demanding the alignment of syllabic prominence and musical metrical strength. Furthermore, constraints concerning musical-metrical strength are less easily violated than constraints demanding the harmonic alignment of poetic/linguistic prominence and note value. The correspondence of poetic/linguistic prominence and tone pitch plays an even less important role in our sample of songs. Overall, the study provides solid empirical evidence about general tendencies underlying the text-setting of classical songs, and for the existence of particular type-setting grammars.

1 Introduction

Linguists and musicologists share a persistent interest in text-setting, that is, in how the language of a poem is matched to musical structure (e.g. Halle & Lerdahl 1993, Kiparsky 2006, Hayes 2009). One domain of investigation concerns the metrical organization of language as reflected in stressed and unstressed syllables, and how this organization is mapped onto various aspects of musical organization, such as musical meter, musical phrasing, note value, tone pitch etc. For example, it is generally assumed that in English vocal music, stressed syllables tend to fall on relatively strong beats of the musical meter (sometimes also called 'stress-matching principle', Palmer and Kelly 1992, Halle and Lerdahl 1993, Proto 2015). However, there is surprisingly little quantitative evidence available for the stress-matching principle (cf. Temperley & Temperley 2013: 521).

With regard to the language side of this matter there is a further complication. Song lyrics themselves are often metrical texts, i.e. their prosodic organization is governed by particular principles that may govern or restrict possible linguistic structures, such as the rhythmic organization of words in a line. The role of the poetic metrical structure in text-setting has remained under-researched so that it is unclear whether the actual stresses or the metrical template are more important for the composer.

In this paper we present a systematic quantitative investigation of how poetic meter (i.e. S and W positions of a metrical template) and syllabic prominence is matched to three dimensions of musical structure: musical meter, note value, and tone pitch. This means that we will not only look at the mapping of stress and musical meter, but also at the mapping of poetic meter and musical meter. And we will look at two important acoustic correlates of stress in English, duration and pitch. These correlates have straightforward correspondences in music, note value and tone pitch.

To assess the range of variability possible in such mappings we created a corpus which comprised three different *Lieder* by the same composer (Haydn, based on three poems by Hunter) and three different settings of the same poem (*Annabel Lee* by Poe), with 161 lines overall.

The empirical results show very clear and statistically significant tendencies in the mapping of language and music. There are, however, also differences between songs and between composers.

To get a better understanding of the observed patterns of mapping we propose an innovative approach in which we use violable constraints and maximum entropy grammars (henceforth 'maxent' grammars) to model the well-formedness of particular mappings as well as the song-specific grammar that underlies the composer's choices in the matching of language and music. This approach is inspired on the one hand by work in generative metrics (Halle & Keyser 1966, 1971, et seq.) and work on metrical well-formedness using maxent grammars (e.g. Hayes et al.

2012), and on the other hand by work in Optimality Theory (Prince & Smolensky 1993) and its application to text-setting (e.g. Hayes 2009).

We develop a set of constraints that evaluate the mapping of poetic and linguistic structure at the one end, and musical structure at the other. Using the maxent approach, we are able to establish constraint hierarchies (i.e. grammars) that underlie particular text-settings. The songs in our sample vary systematically in the probabilities with which particular constraints are violated.

Overall, our study provides solid empirical evidence about general tendencies underlying the text-setting of classical songs, and for the existence of particular type-setting strategies, which can be formalized as grammars.

In the following section, we will introduce in more detail the problems involved in the investigation of the text-setting and lay out our research questions and research strategy. Section 3 explains our methodology, which is followed by the results of the quantitative analysis in section 4. Section 5 presents the background for the constraint-based maxent analysis, and section 6 contains the results thereof. Section 7 discusses our findings.

2 Text-setting: Poetic meter, linguistic rhythm and musical organization

The general problem in text-setting is how linguistic properties and musical properties are mapped onto each other. There is a considerable amount of literature on the topic, but a number of questions have remained unanswered. In this paper we focus on the question of how the rhythmic structure of the text is mapped onto certain aspects of the music, musical meter, note value and tone pitch. With regard to stress and musical meter in English vocal music, previous research is characterized by the general assumption that stressed syllables tend to fall on relatively strong beats of the musical meter (e.g. Palmer and Kelly, 1992, Halle and Lerdahl, 1993). However, surprisingly little quantitative evidence for this mapping can be found in studies of text-setting (cf. Temperley & Temperley 2013: 521).

In addition to stress, there is often another, more abstract level of metrical organization, the poetic meter. Many song lyrics are actually metrical texts themselves, and there is even a whole musical genre, the classical *Lied*, which takes poems, i.e. (in their majority) *metrical* texts, and adds music to them. It is well-known, however, that there is often no one-to-one mapping of strong positions in a poetic-metrical template and stressed syllables. This is illustrated in table 1 with a well-known line from Shakespeare’ sonnet 18. The first line gives the syllables of the words of the line, and the second gives the weak and strong positions of the iambic pentameter template (see Fabb 1997: chapter 2 for an introduction). The third line gives the strength of the stress, as coded by Hayes et al. 2012). ‘3’ and ‘4’ indicate stressed syllables, ‘1’ an unstressed syllable.

Table 1: Example of the mapping of stress and metrical positions in iambic pentameter

syllables	Shall	I	com	**pare**	thee	to	a	**sum**	mer's	day?
template	(W	S)	(W	S)	(W	S)	(W	S)	(W	S)
stress	1	1	1	3	1	1	1	3	1	4

One can easily see that strong positions in the template (‘S’ positions) are not always matched with stressed syllables. Conversely, the stressed syllables of polysyllabic words and the strong phrasal stress at the end of the line are consistently matched with S positions. The apparent variability in the mapping of stress and metrical position raises the question whether it is the actual stresses or the strong and weak positions of the metrical template that are more important for the mapping of musical beats. Depending on whether the composer orients themself towards the stresses or towards the S positions in the template, different matchings and, consequently, different musical patterns might emerge. In the words of Fabb

> it is an empirical question in any particular case whether the matching of text to music is a matching of prosodic phonological structure directly to musical structure, or whether it is mediated through the metrical structure of the text.
>
> (Fabb 1997:105f)

To our knowledge there is no quantitative empirical research available that has explicitly tackled this question. It is therefore one of our aims to clarify for a set of *Lieder* whether it is the stresses or the metrical template that is more predictive of the mapping of beats to the metrical text.

Apart from musical meter there are many other dimensions in music that may stand in a non-arbitrary relationship to the linguistic properties of the text. We have selected two parameters that seem to constitute important properties of any acoustic signal and that are shared by music and language, i.e. duration and pitch (F0 in language). These two properties of sounds are particularly interesting when investigating text-setting because they are also well-known correlates of stress (e.g. Giegerich 1992:179, Hammond 1999:151).[1]

[1] There are also other properties that have been shown to be acoustic correlates of stress in English, for example loudness/intensity, vowel quality and spectral balance (Fry 1955, 1958, Sluyter & van Heuven 1996a, 1996b, Plag et al. 2011). We do not consider these here. Loudness has an obvious correlate in the dynamics of music but is comparatively rarely explicitly used by composers within one song.

Given that in language stressed syllables tend to have higher pitch and are longer than surrounding syllables, one could venture the hypothesis that more prominent syllables in languages have a tendency to be matched with relatively longer and higher tones, in addition to being in positions with stronger beats. For instance, Rodríguez-Vázquez (2010: 251) points out that there are what she calls ‘mismatches’, in which short syllables end up in positions with a strong beat. Hayes (2009) observes that rhythmically strong units tend to be long. Based on these considerations we can make the following empirical predictions concerning the matching of beats, note value and tone pitch on the one hand and stress and templatic positions on the other:

(1) Stress predictions

a. Stressed syllables have a tendency to occur in positions with stronger beats

b. Stressed syllables have a tendency to have longer notes than unstressed syllables

c. Stressed syllables have a tendency to have higher pitch than unstressed syllables

(2) Templatic predictions

a. Syllables in templatic S positions have a tendency to occur in positions with stronger beats

b. Syllables in templatic S positions have a tendency to have longer notes than syllables in templatic W positions

c. Syllables in templatic S positions have a tendency to have higher pitch than unstressed syllables

There is only scattered empirical work on the role of duration in the text-setting of English songs. Hayes & Kaun (1996) find for sung and chanted folksong lines that the number of beats allotted to a syllable corresponds to the syllable’s duration in speech. In environments where one syllable can be matched to two beats, this tends to happen with longer stressed syllables (e.g. *town*), but not with shorter stressed syllables (e.g. *ci* in *city*). Conversely, the Scotch Snap (i.e. a quarter tone tactus divided 1:3 over two syllables, with the first syllable being the stressed one, the second being unstressed) tends to strongly prefer short stressed syllables (see also Temperley & Temperley 2011). Studies focussing on pitch are usually carried out in connection to tone languages (Proto 2015: 120), and little is known with regard to the role of pitch in English text-setting.

The present paper will test the predictions given in (1) and (2) by implementing a quantitative analysis of a corpus of six *Lieder*, three of Joseph Haydn’s English canzonets (1794-95), which are text-settings of poems by the London-based poet Anne Hunter, and three rather recent settings of Edgar Allan Poe's poem *Annabel Lee* (Poe 1997:90), by the composers Scott Gendel, Joel B. New and Jeffrey Quick. The methodology for this empirical investigation will be described in the next section, the results will be presented in section 4.

Given the non-arbitrary distributions of linguistic and musical properties in text-settings, the obvious question is what mechanisms underlie these distributions. We will approach this question by assuming that text-setting is governed by general constraints on the mapping of linguistic and musical structure. The relative importance of these constraints, i.e. their ranking, determines how exactly the mapping plays out. In essence, we will show that certain aspects of a particular text-setting can be described by an underlying constraint hierarchy, the composer's grammar used for the given song.

A similar approach can be found in earlier work by Hayes (e.g. 2009), Rodríguez-Vázquez (2010, 2011) or Keshet (2006). We differ from previous work in three respects. First, we use different constraints, and constraint formulations, in an attempt to capture more adequately the nature of the data. Second, we model not only the mapping of linguistic stress levels and music, but also of the templatic positions and music. Third, we use maximum entropy grammar (e.g. Hayes et al. 2012) to model the well-formedness of particular text-settings.

3 Methodology

3.1 The Corpus

Our corpus consists of six songs varying along various dimensions. Three of them are 18th century songs by the same composer, Joseph Haydn, based on three poems by London-based poet Anne Hunter, *The Mermaid's Song*, *Fidelity*, and *The Wanderer*. The other three songs analysed are more recent (end of 19th/beginning of 20th century) renditions of the 18th century poem *Annabel Lee* by Edgar Allan Poe by three different composers, Scott Gendel, Joel B. New, and Jeffrey Quick.

The choice of these songs is chiefly motivated by our intention to study certain kinds of variation in the mapping of language and music. The three songs based on Hunter's poetry allow for making generalisations about one composer's choices in text-setting across various compositions. The three renditions of *Annabel Lee*, in contrast, allow for comparing text-setting preferences among different composers working with the same text. Finally, the poems show variation in their metrical templates. Table 2 gives an overview of the corpus.

The first row of Table 2 gives the type of feet that we found in the poems. To arrive at these kinds of feet, a metrical template was created for each line of each poem, analogous to Fabb's

(1997) account of iambic pentameter. The metrical template is a series of strong (S) and weak (W) metrical positions to which syllables or feet are matched, and which are grouped into feet (e.g. Fabb 1997: chapter 2, Hayes 1989:221). Each metrical template comes with instructions (i.e. a set of rules, or a kind of grammar) that regulate how the language of the poem is matched to the template. For instance, one of the most important rules Fabb (1997: 41) finds for Shakespeare's iambic pentameter is that the stressed syllables of polysyllabic words must match a strong (i.e. S) position.

Table 2 Overview of the corpus

Song	**Mermaid**	**Wanderer**	**Fidelity**	**Gendel**	**New**	**Quick**
foot types in template	WS	WS, WWS	WS	WS,WWS (highly variable)		
time signature	2/4	3/4	4/4	4/4, 6/4, 3/4	2/2	9/8, 6/8, 3/4
# of syllables	159	130	219	377	377	384
# of S positions	91	47	109	138	138	140
# of W positions	68	79	110	230	230	235

We devised a template for each line of the four poems, following established principles and procedures. The aim was to establish templates that best describe all lines of a poem and are as consistent and uniform as possible across one poem. Feet were matched to lexical and phrasal stresses, minimizing the number of different foot types (trochee, iamb, dactyl, anapaest), the number of extrametrical syllables, and the number of degenerate feet.

We found highly regular and consistent metrical templates for Haydn's three canzonets, whereas for *Annabel Lee* we determined three different templates that in themselves are highly variable. (1) to (4) give the metrical templates for each poem. The parentheses show the grouping of positions into feet (see also Table 3 below). In *Fidelity* shorter lines alternate with longer lines, so that uneven lines have an additional foot of the same kind (WS).

Annabel Lee is highly variable. There are three basic templates, which are given in (4a) to (4c). These templates are variable in themselves in the following fashion. In all three the foot type of the first foot may vary, it can be (S), (WS), (WWS). In (4a) and (4c), if the first foot is (WS) or (WWS), an optional foot may precede, again being either (S), (WS), (WWS). In (4b) and (4c), unfilled weak positions are possible at the edge of phonological phrases.

(1) *The Mermaid's Song*: (S)(WS)(WS(WS)

(2) *Fidelity*: uneven lines: (WS)(WS)(WS)(WS)

even lines: (WS)(WS)(WS)

(3) *The Wanderer*: (WS)(WWS)(WWS)(WWS)

(4) *Annabel Lee*:

a. (WWS)(WWS)

b. (SW)(WS)(WS)(WWS)

c. (SW)(WS)(WS)

In total, 1646 syllables were analysed, and 952 W and 663 S positions are found.

3.2 Coding

For each syllable a number of variables was coded to capture the parameters under investigation, i.e. stress, templatic position, musical metrical strength, note value, and tone pitch. The final coding is shown exemplarily for the first line from Anne Hunter's *The Mermaid's Song* in Table 3 below.

The templatic positions are coded according to the metrical templates previously determined. Extrametrical syllables or the second of two syllables sharing one position were coded as "na" and are therefore excluded from the analysis (31 syllables in total). The positions are grouped into feet illustrated by the brackets in the second line of the table. Stress was coded using four levels of stress according to the criteria developed in Hayes et al. (2012). For the subsequent quantitative analysis of the mapping of stress and beats the four levels were turned into a binary variable with levels 1 and 2 recoded as 'unstressed', and levels 3 and 4 recoded as 'stressed'. This allows us to compare the mapping of metrical positions (which have two values) onto musical dimensions with the mapping of stress onto these dimensions.

Musical metrical strength was assigned to the notes matching a syllable using Temperley & Temperley's (2013) method. Note value was calculated in proportion to the musical bar. For instance, a quarter note in a bar written in four quarter time signature would be assigned the value 0.25. For relative tone pitch we used three different variables. We looked at tone pitch from the preceding note to the note in question and from the note in question to the following one. Here, we coded "stay", "up", and "down" respectively. In addition, we calculated melodic peak scores (Kabak & Domene Moreno 2017) that assign a relative pitch value to notes considering the preceding and the following note at the same time.

In those cases in which one syllable is assigned more than one note, the coding of length and pitch runs into problems. With regard to length we adopted the strategy to add up the proportional durations of all notes assigned to the syllable in question. With regard to pitch we coded the pitch of the first note assigned to the syllable in question.

Table 3: Coding for the first line of *The Mermaid's Song*

syllables	***now***	***the***	***dan-***	***cing***	***sun***	***beams***	***play***
template / feet	(S)	(W	S)	(W	S)	(W	S)
stress level	3	1	3	1	3	2	3
stress (binary)	yes	no	yes	no	yes	no	yes
metrical strength	3	2	3	1	3	2	3
note value	0.5	0.5	0.75	0.25	0.5	0.5	0.75
pitch into	na	up	up	up	down	down	down
pitch out of	up	up	up	down	down	down	na
melodic peak score	na	4	4	6	4	4	na

When a part of the poem is repeated in the musical composition, a syllable or a sequence of syllables from the poems may have more than one musical realisation. In this case, the respective syllables are coded for their textual and musical characteristics for all of their musical representations. If the repetition, however, is an exact repetition of the first instance the syllable occurs in the musical representation, then it is not coded again.[2] Figure 1 below, showing bar lines 80-83 in *Fidelity*, illustrates this procedure. The first repetition of *is cast for me* is annotated again, because the word *is* is assigned a different note. The third repetition of the phrase, however, is an exact repetition if the second and is not annotated again.

Figure 1: Bars 80-83 from *Fidelity* with repetitions of *is cast for me*

4 Results: The mapping of language and music

We systematically tested the mapping of the musical dimensions and templatic positions, and the mapping of the musical dimensions and stress, using statistical tools provided by the statistical analysis software R (R Develoment Core Team 2014). The results of the two kinds of

[2] The examples from the music given in the present writing are our own transcriptions created with the notation software MuseScore, free for download on the respective website.

analyses are very similar, and, for reasons of space, we restrict the presentation of our quantitative results to the mapping of music and templatic positions (sections 4.1 to 4.3). In section 4.4 we will compare the matchings involving the positions in the metrical template with the matchings involving stress. This will answer the question whether composers are guided more by the stresses or by the metrical template.

For each mapping we also fitted a regression model that predicts the value of the musical dimension based on templatic position and stress, respectively. The aim of the regression analyses is twofold. First, it will allow us to assess whether the tendencies (if any) found in these distributional patterns are statistically significant (i.e. not due to random variation). Second, it enables us to reliably compare the predictive power of the two kinds of mappings. The results of all regression models are summarized in section 4.4.

4.1 Musical metrical strength

The distributions of musical metrical strength by metrical position are given in Figure 2. The *x*-axes show the values for musical metrical strength. The *y*-axes give the proportion of metrical positions.

The results meet expectations in all songs (see again prediction (2a) above). Especially Haydn's three canzonets show almost categorical distributions. In *The Mermaid's Song* metrical strength 1 is reserved for W positions, and strength 3 for S positions. Strength 2 can be frequently found in both W and S positions. In *The Wanderer* S positions are almost exclusively paired with the highest metrical strength, and W positions with strength 1 and 2. The difference in distribution between *The Mermaid's Song* and *The Wanderer* may arise from the different time signatures, which is binary in *The Mermaid's Song* and ternary in *The Wanderer*. *Fidelity* has four levels of musical metrical strength and there is a categorical binary distribution of W positions on levels 1 and 2 of musical metrical strength and of S positions on levels 3 and 4 of musical metrical strength.

Gendel's and Quick's renditions of *Annabel Lee* show the consistent tendency of increasing proportions of S positions with rising levels of musical metrical strength. In New's rendition this tendency does not hold for levels 1 and 2 of metrical strength. Overall, the three composers of *Annabel Lee* allow for more variation in the mapping than Haydn.

The regression models for all compositions show that the observed tendencies are significant (see section 4.4, Table 4 for a more detailed documentation of the regression results). This strongly supports the stress-matching principle.

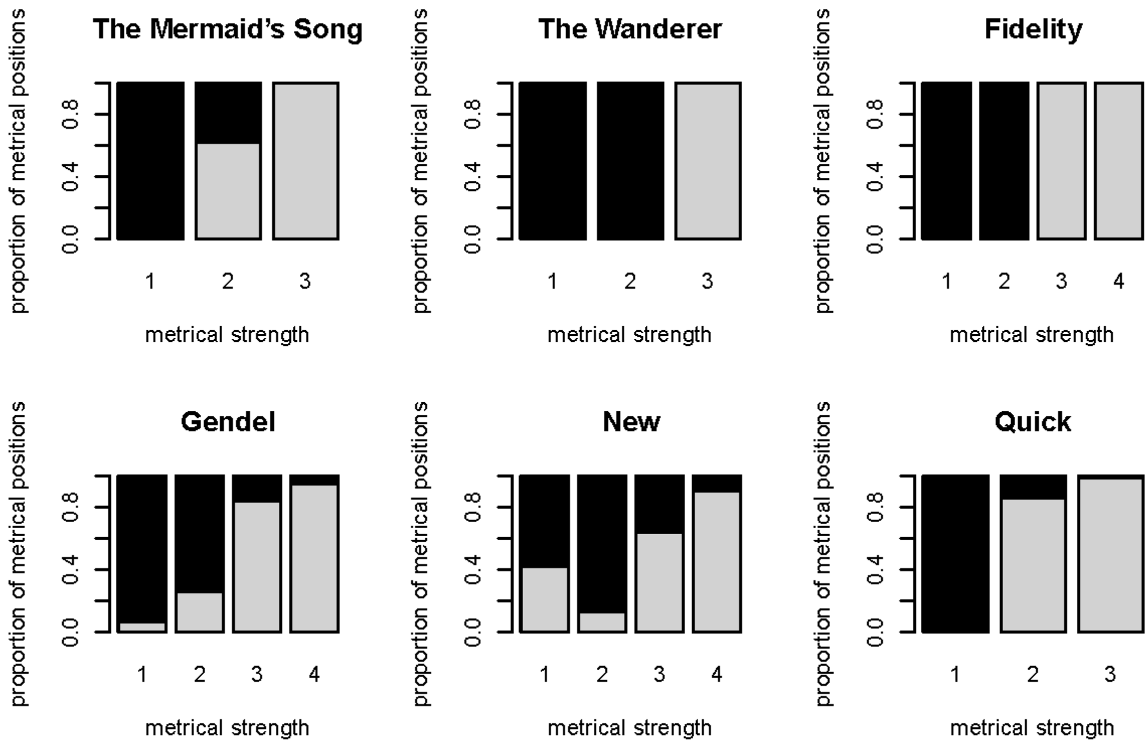

Figure 2: Metrical strength by metrical positions. Proportions for W positions are shown in black, S positions in grey.

4.2 Note value

The second parameter under investigation is note value. Recall that our coding annotates for each syllable whether the note on this syllable is equal, longer or shorter than the preceding note. The results of this analysis are presented in Figure 3.

In all songs, the results largely mirror what has been found for the mapping onto musical metrical strength. Stronger musical metrical strength and S positions tend to coincide. That is, most S positions have longer notes, and longer notes tend to be reserved for syllables in S positions. W positions are mostly reserved to notes of equal or shorter durations and notes with shorter duration overwhelmingly are in syllables in W positions. Haydn's songs are more categorical in their distributions of W and S metrical positions, and the three renditions of *Annabel Lee* again show more variation.

The regression models for all compositions show that the observed tendencies are significant (see section 4.4, Table 4 for more detailed documentation of the regression results). Overall, the results demonstrate that the mapping of note value and metrical positions is an important factor in the text-settings of all of the poems examined here.

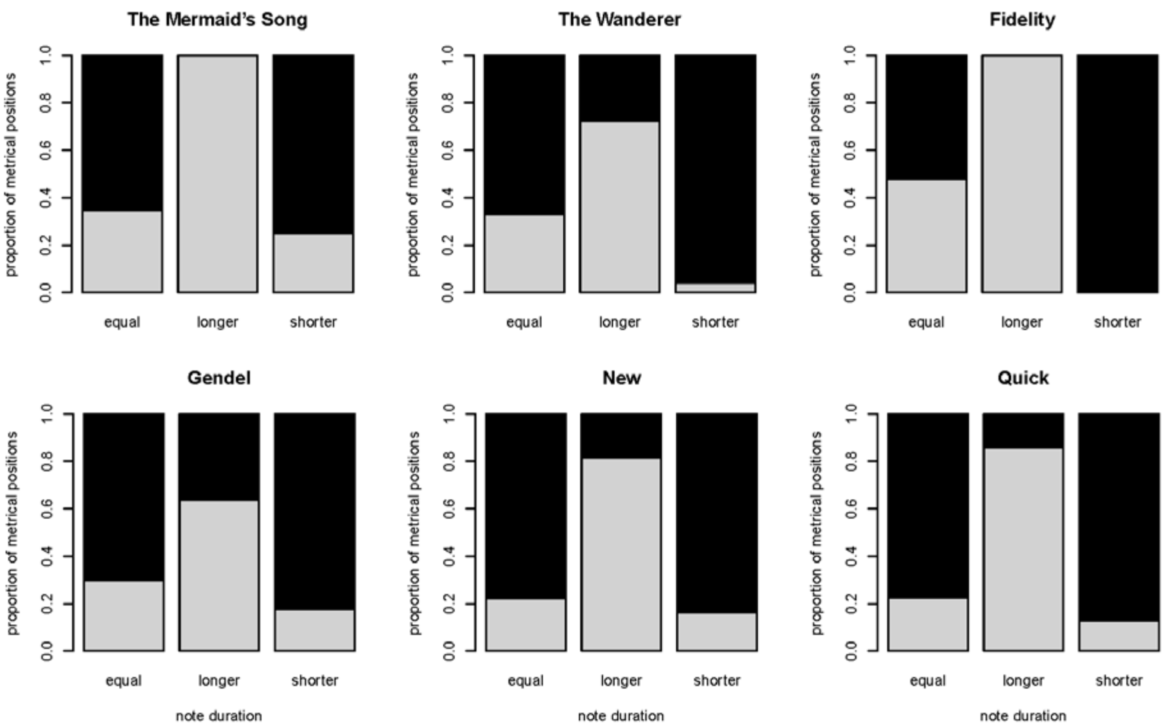

Figure 3: Note duration by metrical positions. Proportions for W positions are shown in black, S positions in grey.

4.3 Tone pitch

For the quantification of tone pitch we used three different variables. The first encodes pitch movement from the previous note into the note in question, the second the pitch movement from the note in question into the following note. These two variables will be used in the maxent analysis in section 5. The third variable, the so-called peak score (Kabak & Domene Moreno 2017), gives the relative prominence of a syllable and is computed on the basis of the first two variables: If the note in question is lower than a preceding (or following, respectively) note it receives the score of 1, a level movement scores 2, and if the note is question is higher, it receives a score of 3. The peak score thus ranges from 2 to 6, with 6 indicating a melodic peak with lower notes on both sides, and 2 a melodic trough with two higher notes on each side. The results for peak scores are presented in Figure 4.

The distributions for the mapping of tone pitch and metrical position show much more random variation than the mappings discussed above. Only Quick's composition shows a strong monotonous increase in the proportions of S positions with increasing prominence, while all other compositions show a more variegated picture with relatively high proportions of S positions also for low prominence values. In the regression analyses, *Quick's* composition shows a strong significant effect, and *The Mermaid's Song, The Wanderer* and *Gendel* very weak significant effects. There are no significant effects of templatic positions on peak scores for *Fidelity* and *New's* setting (see section 4.4, Table 4 for a more detailed documentation of the regression results). Overall, tone pitch seems to play a much less important role in the mapping of language and music than the other two dimensions.

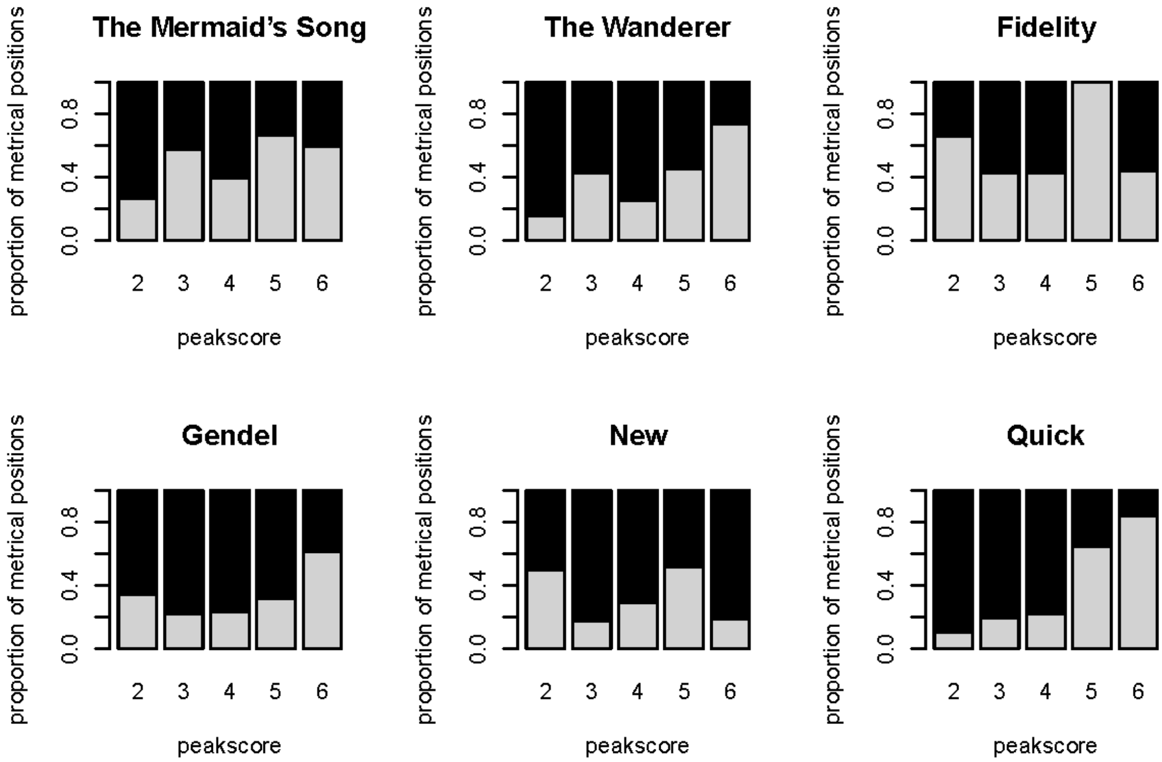

Figure 4: Peak scores by metrical positions. Proportions for W positions are shown in black, S positions in grey.

4.4 Metrical positions vs. stresses

In order to investigate the question whether a composer relies rather on linguistic stresses or on the meter of the text, we use the results of the regression models. The models are documented with their adjusted r^2 values and AICs in table 4. AIC values are shaded. The table shows blank cells for those models where there was no significant effect of templatic position or stress. For the comparison of unnested models the AIC is a standard measure. A lower AIC indicates a better model fit.

Before turning to the model comparisons, let us have a look at the strength of the effects, as gauged by the r^2 values. If we compare them across the different compositions and across the three musical dimensions, we can easily see that, generally, there are strong effect sizes for metrical strength (with New as an exception), somewhat weaker effects for note value, and extremely weak (or no) effects for tone pitch (with the exception of Quick).

The comparison of the AICs for the template-based mapping with those of the stress-based mapping reveals higher AIC's with the mapping of stresses. The only exception is the model fitted to tone pitch for *The Mermaid's Song*. In this model, we have the same AIC for both models, but both models have a very poor fit (r^2 = 0.03 for the template, 0.02 for the stresses).

In sum, our data provide evidence that there is a greater tendency among the composers in our sample to match musical properties to templatic positions rather than to stressed and unstressed syllables.

Table 4: R^2 and AIC values of regression models for the mapping of metrical strength, note value and tone pitch onto language, based on templatic positions and stress, respectively. 'M'= *The Mermaid's Song*, 'W' = *The Wanderer*, 'F' = *Fidelity*, 'G' = Gendel, 'N' = New, 'Q' = Quick.

	M		**W**		**F**		**G**		**N**		**Q**	
	r^2	**AIC**	**r^2**	**AIC**	**r^2**	**AIC**	**r^2**	**AIC**	**r^2**	**AIC**	**r^2**	**AIC**
metrical strength												
template-based	0.76	183	0.74	94	0.82	213	0.59	856	0.03	970	0.79	355
stress-based	0.52	294	0.68	120	0.39	476	0.55	906	0.04	987	0.73	452
note value												
template-based	0.32	255	0.24	228	0.22	225	0.18	809	0.28	605	0.32	601
stress-based	0.26	267	0.22	233	0.10	255	0.18	827	0.24	641	0.34	617
note pitch												
template-based	0.03	439	0.10	377			0.04	957			0.23	877
stress-based	0.02	439	0.09	396			0.04	974			0.20	897

5 Constraint-based analysis: A maxent grammar of text-setting

5.1 The constraints

For our constraint-based analysis we will use constraints that address three different mapping problems: stress matching, bracket matching and contextual salience. We will discuss each in turn.

Stress matching constraints have played an important role in previous work (e.g. Hayes 2009) as they are needed to account for the tendency of stressed syllables to coincide with strong, rather than weak, beats. We will therefore implement both the matching of stress and of templatic positions to musical meter.

Hayes (2009) introduces two kinds of stress matching constraints, targeting either stressed or unstressed syllables. These constraints penalize stressed syllables in musical metrical positions that are of medium or weak strength, and they penalize unstressed syllables in strong and in weak musical metrical positions.

We differ from Hayes (2009) by taking a different direction in the coding of the constraints and by parametrizing the stress matching constraints more systematically. Given that a composer needs to map music onto an existing text, we feel that constraints regulating this mapping should be conceptualized as working in this direction, too. We therefore propose the two constraint families given in (3) and (4) for metrical positions, and the ones given in (5) and (6) for stresses. 'W' and 'S' refer to the strong and weak positions in the poetic metrical template, 'beat' refers to the positions in the musical meter. The constraint in (3) penalizes beats in W positions. The harmonic hierarchy given in the second line of (3) ensures that weaker beats in W positions are more easily tolerable than stringer beats. The constraint in (4) takes care that strong beats are placed on S, with strongest beats being most desirable in this position. The constraints in (5) and (6) work analogously for the matching of stressed and unstressed syllables and beats.

(3) *BEAT IN W: Avoid beats in W positions.
*BEAT4 IN W >> *BEAT3 IN W >> *BEAT2 IN W >> *BEAT1 IN W

(4) *BEATLESS IN S: Avoid having no beats in S positions.
*BEAT1 IN S >> *BEAT2 IN S >> *BEAT3 IN S >> *BEAT4 IN S

(5) *BEAT ON UNSTRESSED: Avoid beats on stressless syllables.
*BEAT4 ON UNSTRESSED >> *BEAT3 ON UNSTRESSED >> *BEAT2 ON UNSTRESSED >>
*BEAT1 ON UNSTRESSED

(6) *BEATLESS ON STRESSED: Avoid having no beats on stressed syllables.
*BEAT1 ON STRESSED >> *BEAT2 ON STRESSED >> *BEAT3 ON STRESSED >> *BEAT4 ON STRESSED

For reasons of space, we restrict our constraint-based analyses of note value and pitch to the mapping of these parameters onto metrical positions. With regard to note value, the most harmonic ranking would ensure that notes are longer in S positions than in surrounding W positions. This is the work of the constraints put forward in (7) and (8), which are inspired by the constraints used in Hayes et al. (2012) to model the mapping of stress and metrical templatic positions in Shakespeare's sonnets and Milton's *Paradise Lost*. The constraints in (7b, d) and (8b, d) are stricter variants of the length constraints, as they also penalize notes of the same length, and not only notes that are longer.

(7) a. *LONGER AFTER S: Avoid longer notes after S
b. *LONGER/EQUAL AFTER S: Avoid longer notes, or notes of equal length, after S
c. *LONGER BEFORE S: Avoid longer notes before S
d. *LONGER/EQUAL BEFORE S: Avoid longer notes, or notes of the same length, before S

(8) a. *SHORTER AFTER W: Avoid shorter notes after W
b. * SHORTER/EQUAL AFTER W: Avoid shorter notes, or notes of the same length, after S
c. *SHORTER BEFORE W: Avoid shorter notes, or notes of equal length, before W
d. *SHORTER/EQUAL BEFORE W: Avoid shorter notes, or notes of equal length, before W

Similar constraints hold for tone pitch, as given in (9) and (10).

(7) a. *HIGHER PITCH AFTER S: Avoid higher pitch after S
b. *HIGHER/SAME PITCH AFTER S: Avoid higher pitch, or pitch of the same height, after S
c. *HIGHER PITCH BEFORE S: Avoid higher pitch before S
d. *HIGHER/SAME BEFORE S: Avoid higher pitch, or pitch of the same height, before S

(8) a. *LOWER PITCH AFTER W: Avoid lower pitch after W
b. *LOWER/SAME PITCH AFTER W: Avoid lower pitch, or pitch of the same height, after W
c. *LOWER PITCH BEFORE W: Avoid lower pitch before W
d. *LOWER/SAME PITCH BEFORE W: Avoid higher pitch, or pitch of the same height, before S

Let us now turn to the kind of grammar that will be used to model the text-setting.

5.2 Maximum entropy grammar

Maxent grammars emerge as the result of a learning algorithm that makes probabilistic generalizations, balancing accuracy and generality. The maximum entropy formalism has a long tradition in machine learning literature (e.g. Jaynes 1957, Berger et al. 1996) and has recently been used in linguistics, for example in studies of phonotactics and metrics (e.g. Goldwater & Johnson 2003, Wilson 2006, Hayes & Wilson 2008, Martin 2011, Hayes et al. 2012).

A maxent grammar contains a list of numerically weighted constraints. This is different from classical Optimality, in which constraints are strictly ranked, leaving no room for variability and probability. The weight of the constraint in a maxent grammar represents its strength, which is a function of the number of times it is violated. The more violations, the lower the weight of the constraint. The constraints and their weights constitute the grammar. Given the constraints and their weight, the probability of particular potential output structures can be computed. This probability indexes the well-formedness of the structure in question.

For the implementation we make use of the maxent grammar tool hosted by Bruce Hayes (available at https://linguistics.ucla.edu/people/hayes/MaxentGrammarTool). Technical details about the algorithm and its implementation can be found in the literature referenced above and will not be discussed in detail here. In essence, the algorithm is fed a list of constraints and the number of violations of each constraint for each candidate. Based on the violations the algorithm computes the constraint weights.[3] In the next section, we will present the results of this implementation.

6 Results: Constraint rankings

6.1 Musical metrical strength

The constraint weights for the mapping of musical meter and templatic positions is given in Table 5, the weights for the mapping of musical meter and stress is given in table 6.

[3] For the computation of the constraint weights, it is possible to set a specific Gaussian prior for each constraint to penalize high weights. We chose $\mu=1$ and $\sigma^2=5.0E8$ for all constraints in all computations to minimize this penalization, and to obtain weights at a magnitude that makes them easy to compare.

Table 5: Constraint weights for the mapping of musical meter and templatic positions

	Mermaid	**Wanderer**	**Fidelity**	**Gendel**	**New**	**Quick**
*BEAT1 IN S	5.00	5.00	5.00	6.60	1.25	5.72
*BEAT2 IN S	0.00	0.00	5.00	4.95	5.68	0.58
*BEAT3 IN S	0.00	0.00	0.00	1.93	6.03	0.33
*BEAT4 IN S			0.00	1.15	3.45	
*BEAT4 IN W			5.00	6.40	6.24	
*BEAT3 IN W	5.00	5.00	5.00	3.20	2.08	5.56
*BEAT2 IN W	0.00	0.00	0.00	0.56	0.17	0.70
*BEAT1 IN W	0.00	0.00	0.00	0.15	0.18	0.02

In all but New's text-setting we find harmonic rankings. Haydn has very clear preferences. In a ternary musical rhythm, i.e. one with three levels of strength, S positions strongly disfavor the weakest beat, and in the quarternary rhythm (with four levels of strength) S positions disfavor the two weakest beats. Gendel and Quick show very similar grammars. New's version of *Annabel Lee* shows a harmonic ranking for weak positions but a non-harmonic ranking for S positions, with S positions attracting the weakest beats (instead of the strongest, as in harmonic rankings).

The mapping of stresses, as shown in table 6, shows a similar picture. For stressed syllables all compositions but New's have harmonic rankings, for unstressed syllables *Fidelity* shows a non-harmonic ranking that disfavors the weakest beats and favors the strongest beats on unstressed syllables.

Table 6: Constraint weights for the mapping of musical meter and stress

	Mermaid	**Wanderer**	**Fidelity**	**Gendel**	**New**	**Quick**
*BEAT1 ON STRESSED	5.00	5.00	5.00	6.47	1.25	5.61
*BEAT2 ON STRESSED	0.00	0.00	0.00	5.08	6.07	0.47
*BEAT3 ON STRESSED	0.00	0.00	0.00	1.78	5.69	0.27
*BEAT4 ON STRESSED			0.00	1.06	3.25	
*BEAT4 ON UNSTRESSED			1.45	6.35	6.19	
*BEAT3 ON UNSTRESSED	6.41	5.00	6.36	4.04	1.86	5.85
*BEAT2 ON UNSTRESSED	3.92	0.00	0.31	0.84	0.16	3.90
*BEAT1 ON UNSTRESSED	1.31	0.00	4.54	0.26	0.16	0.15

6.2 Note values

The rankings of the constraints that are responsible for the mapping of note values onto metrical positions are given in table 7.

Table 7: Constraint weights for the mapping of note value and templatic position

	Mermaid	**Wanderer**	**Fidelity**	**Gendel**	**New**	**Quick**
*LONGER AFTER S	5.1	6.0	5.1	5.9	5.5	5.74
*LONGER/EQUAL AFTER S	0.1	1.8	0.1	3.4	0.7	1.08
*LONGER BEFORE S	5.0	5.8	5.1	5.7	5.7	5.49
*LONGER/EQUAL BEFORE S	0.0	1.3	0.1	3.9	1.0	0.62
*SHORTER AFTER W	5.9	5.2	5.0	5.9	5.9	5.77
* SHORTER/EQUAL AFTER W	1.5	0.2	0.0	3.4	1.5	1.15
*SHORTER BEFORE W	5.0	5.8	5.1	6.0	5.6	5.48
*SHORTER/EQUAL BEFORE W	0.0	1.1	0.1	3.2	0.8	0.60

All compositions are based on harmonic rankings that strongly favor S positions to have notes of longer or equal duration as the surrounding positions, and that strongly favor W positions to have notes of shorter or equal duration than its preceding and following position. The weights of the constraints vary only slightly across compositions. Only Gendel’s weights are remarkably different from the others in that there is a stronger tendency for S positions to be occupied by longer notes and for W positions to be occupied by shorter than the surrounding positions.

6.2 Tone pitch

The mapping of pitch and metrical positions is much less harmonic than the previous mappings, as documented in table 8.

The Wanderer and New’s version of *Annabel Lee* show harmonic patterns throughout, with S positions favoring higher pitches and W positions lower pitches than the surrounding positions. The weights of the clearly ranked constraints on the respective hierarchies (e.g. 5.8 and 3.6 in the first two cells for *The Mermaid’s Song*) are, however, closer to each other than the weights in the previously discussed hierarchies for metrical strength and note value (e.g. 5.0 vs. 0.0, or 5.0 vs. 0.1 in the corresponding cells in tables 2, 3, and 4). This means that there is more variation.

Table 8: Constraint weights for the mapping of tone pitch and templatic position

	Mermaid	**Wanderer**	**Fidelity**	**Gendel**	**New**	**Quick**
*HIGHER PITCH AFTER S	5.8	6.0	5.3	5.4	6.0	5.91
*HIGHER/SAME PITCH AFTER S	3.6	2.8	4.6	4.5	2.9	1.58
*HIGHER PITCH BEFORE S	5.4	5.9	5.2	5.8	6.0	5.44
*HIGHER/SAME BEFORE S	4.5	3.6	4.8	3.7	2.8	4.46
*LOWER PITCH AFTER W	5.4	5.9	5.2	5.6	6.0	5.79
*LOWER/SAME PITCH AFTER W	4.5	3.6	4.8	4.1	2.7	3.81
*LOWER PITCH BEFORE W	5.9	6.0	5.3	5.3	6.0	6.03
*LOWER/SAME PITCH BEFORE W	3.4	2.9	4.6	4.7	2.8	2.41

The other four grammars are characterized by a mixture of weak harmonic rankings and near-ties in the constraint rankings. We use the term ‘near-tie’ for weights whose difference is smaller than 1. These are marked in grey in table 5. Although still harmonic the weights in the grey cells are so close to each other that the grammar allows for a great number of violations with both constraints, and hence rather random variation.

Overall, the mapping of pitch and templatic positions is much less constrained than the other mappings investigated in this paper.

7 Discussion and conclusion

The present paper has investigated how poetic meter (as formalized in metrical templates) and syllabic prominence is matched to three dimensions of musical structure: musical meter, note value, and tone pitch. Based on the quantitative analysis of a corpus of three different *Lieder* by Haydn and three different settings of the poem *Annabel Lee*, it was possible to discern very clear tendencies in the mapping of language and music.

The strongest tendency is the matching of prosodically strong elements of the language with strong elements in the musical meter. Another strong tendency is the alignment of prosodically strong elements of the language with longer note values. Generally, there is only a weak tendency to align higher tone pitch with strong prosodic units.

The analysis also revealed that there are clear differences between songs and between composers in how they relate language and music. Composers may choose text-settings that go against the tendencies just described. For instance, New does not adhere well to the stress-matching principle, and Gendel tends to align pitch peaks with strong prosodic positions.

The quantitative analysis also allowed us to investigate a long-standing problem in the setting of metrical texts, i.e. the question of whether it is the metrical template or the stress distribution that is responsible for the mapping of linguistic prominence and musical prominence (as manifested in beats, note length and tone pitch). The comparison of both kinds of mapping using regression models yields a clear picture. There is a tighter mapping observable of musical prominence with the S and W positions of the metrical template of the poem than with the stressed and unstressed syllables in the text.

The mappings as manifested in the six songs under investigation were formally modelled using an innovative approach inspired by Optimality Theory and some recent studies on phonotactics and poetic meter. In this approach, violable constraints that mitigate against certain types of mappings of text and music, are implemented in a maximum entropy grammar. Songs vary in the probabilities with which particular constraints are violated. Using the maxent approach, it is possible to establish constraint hierarchies, i.e. individual grammars, that underlie particular text-settings. This formalization allows us to make verifiable claims about the properties of the songs under consideration, but also about the nature of text-setting.

We found that in all text-settings, there is a strong tendency towards harmonic rankings in which linguistic prominence (be it via the template, or via the stresses) aligns with musical metrical prominence. Furthermore, there is a strong tendency towards harmonic rankings that regulate the mapping of templatic positions and note values. With regard to tone pitch, the alignment is much less harmonic and less strong.

Overall, in the songs investigated in this study, we find that of the three dimensions under investigation that are shared by language and music, i.e. rhythmic prominence, duration and pitch, the first two are much more important than tone pitch in matching linguistic structure and musical dimensions.

Our maxent implementation has also demonstrated that it is possible to devise formal systems, i.e. grammars, that can bring about the particular mappings of language and music that characterize individual compositions. In this way it is possible to pin down a composer's individual strategy of dealing with particular kinds of mappings. For example, a composer may choose disharmonic rankings for some constraints and harmonic rankings for others. In sum, the maxent implementation has provided solid empirical evidence for the existence of particular type-setting strategies. The consequences of such strategies for the character of the songs may be very interesting to look at from a musicological perspective. The grammars devised in this study might thus be used as a springboard for the development of new research questions in this field.

Future work may show whether the results of this study carry over to other genres and composers, or hold also for a larger corpus of *Lieder*, representing other composers.

References

Primary sources

Poe, Edgar Allan. 1997. *Spirits of the dead. Tales and poems*. London: Penguin.

Secondary sources

Berger, Adam L., Vincent J. Della Pietra & Stephen A. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics* 22(1). 39–71.

Fabb, Nigel. 1997. *Linguistics and literature: Language in the verbal arts of the world*. Oxford: Blackwell.

Fry, Dennis B. 1955. Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America* 27. 765–768.

Fry, Dennis B. 1958. Experiments in the perception of stress. *Language and Speech* 1. 126–152.

Giegerich, Heinz J. 1992. *English phonology: An introduction* (Cambridge textbooks in linguistics). Cambridge and New York: Cambridge University Press.

Goldwater, Sharon & Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.), *Proceedings of the Stockholm workshop on variation within Optimality Theory*, 111–120. Stockholm: University of Stockholm.

Halle, Morris & Samuel J. Keyser. 1966. Chaucer and the study of prosody. *College English* 28(3). 187–219.

Halle, Morris & Samuel J. Keyser. 1971. *English stress: Its form, its growth, and its role in verse*. New York: Harper and Row.

Halle, John & Fred Lerdahl. 1993. A generative text-setting model. *Current Musicology* 55(3-23).

Hammond, Michael. 1999. *The phonology of English: A prosodic optimality-theoretic approach.* Oxford: Oxford University Press.

Hayes, Bruce. 1989. The prosodic hierarchy in meter. In Paul Kiparsky & Gilbert Youmans (eds.), *Rhythm and meter*, 201–260. Orlando: Academic Press.

Hayes, Bruce. 2009. Text-setting as constraint conflict. In Jean-Louis Aroui & Andy Arleo (eds.), *Towards a typology of poetic forms: From language to metrics and beyond*, 43–52:

John Benjamins.

Hayes, Bruce & Abigail Kaun. 1996. The role of phonological phrasing in sung and chanted verse. *The Linguistic Review* 13(3-4). 243–304.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39(3). 379–440.

Hayes, Bruce, Colin Wilson & Anne Shisko. 2012. Maxent grammars for the metrics of Shakespeare and Milton. *Language* 88(4). 691–731.

Jaynes, Edwin T. 1957. Information theory and statistical mechanics. *Physical Review* 106(4). 620.

Keshet, Ezra. 2005. Relatively optimal text-setting. Ms. Cambridge: MIT.

Kiparsky, Paul. 2006. A modular metrics for folk verse. In B. E. Dresher & Nila Friedberg (eds.), *Formal Approaches to Poetry*, 7–49. Berlin/New York: Mouton de Gruyter.

Martin, Andrew. 2011. Grammars leak: Modeling how phonotactic generalizations interact within the grammar. *Language* 87(4). 751–770.

Palmer, Caroline & Michael H. Kelly. 1992. Linguistic prosody and musical meter in song. *Journal of Memory and Language* 31(4). 525.

Plag, Ingo, Gero Kunter & Mareile Schramm. 2011. Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics* 39. 362–374.

Prince, Alan & Paul Smolensky. 1993. *Optimality Theory. Constraint Interaction in Generative Grammar*. Ms. New Brunswick, Bolder: Rutgers University; University of Colorado.

Proto, Teresa. 2015. Prosody, melody and rhythm in vocal music: The problem of text-setting in a linguistic perspective. *Linguistics in the Netherlands* 32(1). 116–129.

R Development Core Team. 2014. *R: A Language and Environment for Statistical Computing. Vienna*, Austria: R Foundation for Statistical Computing. http://www.r-project.org.

Rodríguez-Vázquez, Rosalía. 2010. *The rhythm of speech, verse and vocal music: A new theory*. Bern: Peter Lang.

Rodríguez-Vázquez, Rosalia. 2011. Constraint Ranking in English Broadside Ballads. In Christoph Küper (ed.). *Current trends in metrical analysis*. Frankfurt am Main, Berlin, et al.: Peter Lang.

Sluijter, Agaath M.C. & Vincent J. van Heuven. 1996a. Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Proceedings of the 4th International Conference on Spoken Language Processing*, Philadelphia, 630–633.

Sluijter, Agaath M.C. & Vincent J. van Heuven. 1996b. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustic Society of America* 100(4). 2471–2485.

Temperley, Nicholas & David Temperley. 2013. Stress-meter alignment in French vocal music.

Journal of the Acoustical Society of America 134(1). 520–527.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. Cognitive Science 30(5). 945–982.