Reality and perception of stress-beat mismatches in German *Volkslied*

Teresa Proto

[1]Leiden University/Meertens Institute, Netherlands
t.proto@hum.leidenuniv.nl

Abstract

This paper aims to contribute to the study of how prosody interacts with musical structure in songs by providing empirical data for German textsetting. The focus is on the way (primary) stressed syllables in words are aligned with the metrical (musical) grid in *Volkslieder*. Besides sketching a tentative typology of the kinds of prominence mismatches allowed in German, this study also explores the issue of how such mismatches are perceived by native speakers. The results are discussed in the light of a theory of meter and accentuation in music and the nature of stress in language.

Index Terms: Textsetting, German, prominence matching, accentuation, melodic accent

1. Introduction

Previous research on textsetting has revealed a clear-cut distinction between English, on the one hand, and the Romance languages, on the other. It would appear that in languages traditionally classified as syllable-based, setting lexically stressed syllables to a weaker metrical position than the adjacent stressless syllables (in the same word) is commonplace, whereas in a stress-based language like English, this is not allowed. This principle of agreement between linguistic and musical prominence is known as stress-beat matching (henceforth SBM). In the present research, I set out to establish to what extent SBM is enforced in German, another stress-based language.

Stress-beat mismatches have been reported to have a perceptual correlate in Spanish [1], where lexical stress is perceived as having been "shifted" to a neighboring syllable. Similarly, settings that violate prominence matching in English are usually rejected as ill-formed by native speakers by virtue of the resulting fatal distortion of local stress patterns [2], which "places the rhythmic cues of stress on the 'wrong' syllables" [3, p. 292]. This paper thus addresses the issue of whether and to what extent stress-beat mismatches are perceptible to native speakers of German. The results are discussed in the light of a theory of meter and accentuation in music and the nature of stress in language.

2. Prominence matching in German *Volkslied*

This study is based on an analysis of *Volkslieder* drawn from two songbooks [4 and 5] and supplemented by online sources (mostly performances of the songs available on YouTube). In line with previous similar studies [3, 6, 7 and 8], the analysis is confined to simplex words. This excludes, for instance, stress-beat misalignments involving primary stress in compounds. Prominence mismatches involving melismas (i.e. assignments of one syllable to multiple notes) are not considered either. In what follows, mismatch, misalignment, and prominence violation will be used as synonyms.

2.1. Metrical grids

The grid depicted in Fig. 1 will be used throughout this paper to represent the musical settings. This diagram [cf. 9] is a combination of two grids: the metrical-melodic grid, which is above the text, and the prosodic grid, which is below. The upper part of the diagram captures the metrical structure of the tune, conceived of as an alternation of downbeats and upbeats, as well as the sequence of musical pitches (capital letter and number identifying the pitch's octave). The higher the column of x's, the stronger the relative strength of the beat. In the lower part of the diagram, the columns of dots capture the prominence relations within the text: while all syllables are assigned one dot, only syllables carrying primary lexical stress and monosyllabic content words receive two dots. For ease of reference, each metrical position is assigned a number on top of the grid. The diagram allows us to visually compare the metrical structure of the tune with the prosodic structure of the lyrics, and to easily detect differences in terms of their relative prominence. Whenever the stressed syllable of a word is matched to a weaker metrical position than the unstressed syllables, a stress-beat mismatch occurs. In this study, I will only look at prominence matching within simplex polysyllabic words.

Fig. 1: *Metrical-prosodic grid*

It will be shown that although stress-beat mismatches are disallowed in German, they don't fall into a single homogenous class; rather, they can be classified into at least three groups.

2.2. Type 1 mismatch

The setting in Fig. 2 exemplifies a mismatch involving the primary stressed syllable of a simplex content word. At the beginning of the line (positions 2–3), the tonic syllable of the adjective *selig* ("blessed") is matched to a weaker metrical position than the post-tonic syllable *.lig*. Formally, this gives rise to a stress-beat mismatch. One sung rendition of the setting in Fig. 2 was recorded and submitted to three native speakers who were familiar with the singing idiom under study. In the intuitions of the consulted speakers, this

10.21437/SpeechProsody.2018-112

violation—while acceptable—results in the word stress being perceived as having been displaced from the first to the second syllable.

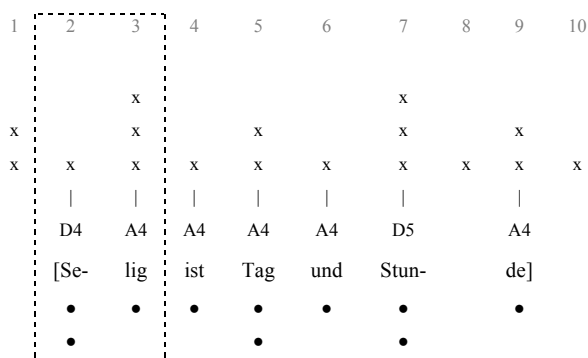| 1	2	3	4	5	6	7	8	9	10
		x				x			
x		x		x		x		x	
x	x	x	x	x	x	x	x	x	x
	|	|	|	|	|	|		|	
	D4	A4	A4	A4	A4	D5		A4	
	[Se-	lig	ist	Tag	und	Stun-		de]	
	•	•	•	•	•	•		•	
	•		•		•				

Fig. 2: *Mismatch in monomorphemic (content) words*

Misalignments of this type are extremely rare in my corpus (four occurrences, repetitions excluded), and they only involve disyllabic words with a trochaic stress pattern. This is similar to English folk song, where the (rare) occurrences of so-called "lexical inversion" concern words with a falling lexical stress contour [8].

2.3. Type 2 mismatch

Fig. 3 exemplifies a mismatch involving words with two possible stress patterns (stress doublets). As in the previous example, this configuration occurs at the beginning of the line.

The syllables matched to positions 1 and 4 are part of the pronominal adverb *danach*, which is morphosyntactically composed of an anaphoric element *da-* and a relational element *-nach*. In everyday speech, this word is stressed on the second syllable if it has temporal function; in all other cases it can be stressed on the anaphoric element for emphasis or focus.
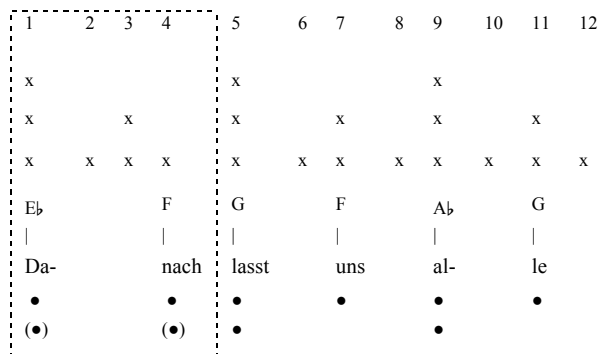
1	2	3	4	5	6	7	8	9	10	11	12
x				x				x			
x		x		x		x		x		x	
x	x	x	x	x	x	x	x	x	x	x	x
E♭		F		G		F		A♭		G	
|		|		|		|		|		|	
Da-		nach		lasst		uns		al-		le	
•		•		•		•		•		•	
(•)		(•)		•				•			

Fig. 3: *SBM and stress doublets*

In German, there is a whole class of adverbs with two alternative stress patterns (*darum*, *trotzdem*, *dazu* etc.). In most cases, they have historically resulted from a compositional process and their internal morphological boundaries are still well recognizable. For this type of word, a mechanical scansion cannot reveal whether stress and (strong) beat are correctly aligned: it would be necessary to look into the information structure of each sentence in which they occur. A similar issue arises with other word classes, such as complex adjectives, for which two readings (and two stress

patterns) are available. Adjectives like *blutarm* and *steinreich*, which syntactically show a head-argument structure, are prosodically ambivalent: they always have initial stress in attributive position, whereas in predicative (prenominal) position, the main stress can be either on either the first syllable (literal meaning) or the second (metaphorical reading) [10, p. 310]. A search performed on a sample of settings revealed that information structure and syntax are largely ignored when stress doublets are set in relation to rhythmic structure in songs.

2.4. Type 3 mismatch

The third and last type is exemplified in Fig. 4. Here the word *Weibsen*, at the end of the line, is involved in a stress-beat mismatch. Unlike prominence violations of type 1 (and 2), this configuration is not perceived as conflicting with the natural stress pattern of the word. Although formally a stress-beat mismatch, it does not result in a distorted perception of the prosodic shape of the word.
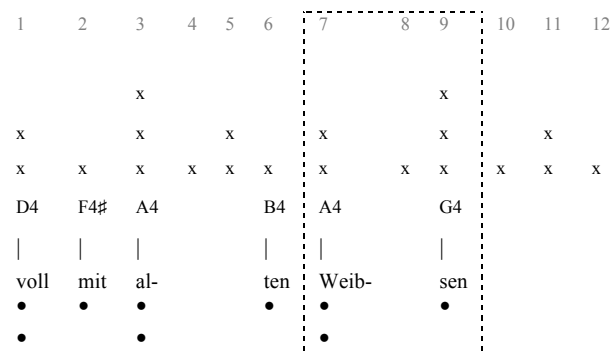
1	2	3	4	5	6	7	8	9	10	11	12
	x							x			
x	x	x				x		x		x	
x	x	x	x	x	x	x		x	x	x	x
D4	F4♯	A4			B4	A4		G4			
|	|	|			|	|		|			
voll	mit	al-			ten	Weib-		sen			
•	•	•			•	•		•			
	•		•			•					

Fig. 4: *Non-perceptible stress-beat mismatch*

2.4.1. Constructs

In order to investigate Type 3 mismatches more closely, a number of constructs were made up based on the original tune in Fig. 4. In each construct, slight ad hoc changes were introduced locally in the penultimate musical bar and up to the mismatch position. Sung renditions of the constructs as well as the original setting were submitted to the informants. It turned out that very simple local manipulations of the melodic and temporal contour were capable of affecting native speakers' judgments of well-formedness. In one of the constructs (Fig. 5), a change was introduced at the melodic level in the correspondence of the syllables *.ten* and *Weib.*, whereby their pitch values were inverted. In the cadence, in place of the B4–A4–G4 transition of the original tune, we introduced the transition A4–B4–G4. Based on the intuitions on my informants, the simple inversion of the second and third pitches in the penultimate bar is sufficient not only to make the misalignment perceptible to their ears, but also to make the setting altogether ill-formed.
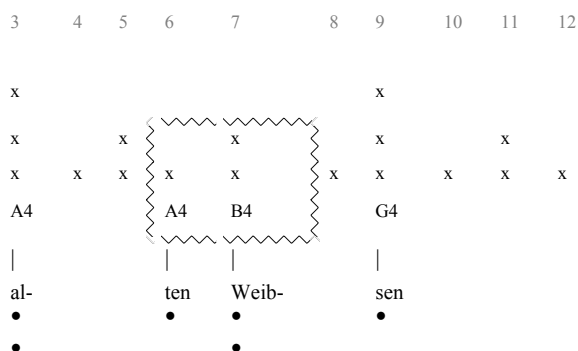
```
3     4    5    6    7        8    9    10    11    12

x                                  x

x          x         x             x         x

x     x    x    x    x        x    x    x    x    x

A4                   A4   B4        G4

|                    |    |         |

al-                  ten  Weib-     sen
●                    ●    ●         ●

●                         ●
```

Fig. 5: *Construct #1*

In another construct (Fig. 6), the duration values of the notes matched to the syllables *al.*, *.ten* and *Weib.* were altered. The original rhythmical formula, consisting of a dotted quarter note followed by an eighth note and a quarter note, was replaced by a formula starting on a half note followed by two eighth notes matching, respectively, the syllables at positions 3, 7 and 8 in Fig. 6 (compare the original tune in Fig. 4). This shifted the onset of the syllable *Weib.* to the weakest beat of the musical bar (= pos. 8), which is in 3/4 time.
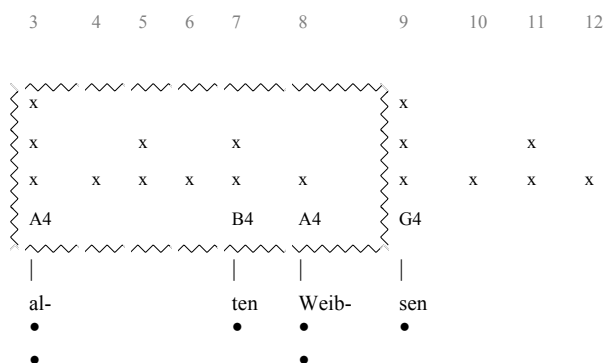
```
3     4    5    6    7        8    9    10    11    12

x                                  x

x          x         x             x         x

x     x    x    x    x        x    x    x    x    x

A4                   B4   A4        G4

|                    |    |         |

al-                  ten  Weib-     sen
●                    ●    ●         ●

●                         ●
```

Fig. 6: *Construct #2*

Construct #2 was judged awkward by the speakers not only as a consequence of the stress-beat mismatch, but also because the syllable *.ten* ended up being scanned as the prosodic head of the following word *Weibsen* instead of grouping together with the preceding syllable *al*. This construct nicely illustrates one of the ways in which musical meter interacts with linguistic constituent structure, and how duration and inter-onset intervals may affect perceived accent in music.

3. Discussion

In German *Volkslieder*, the distribution of misalignments does not appear to be confined to specific parts of the line, or to be restricted to specific parts of speech (function as well as content words may be involved). The data partly supports the claim that in stress-based languages, the principle of SBM is strictly enforced. Like in English, in German the occurrence of prominence violations is highly restricted. Unlike in English, however, a qualitative difference exists in German between misalignments of types 1 and 2, on the one hand, and mismatches of type 3, on the other. What is special about type 3 items is that, although the prosodic structure of the word is formally in conflict with the metrical structure of the tune, this

has no perceptual reality. To my knowledge, stress-beat misalignments of this kind have not been reported in previous literature on textsetting. At present, I have no explanation for this difference; however, several considerations are in order, some of which concern the grammar of the music, others of which have relevance for the linguistic domain.

Let us consider the nature of linguistic stress and the effects of stress violation in German. While the mental reality of prosodic prominence is undisputed, an unambiguous phonetic correlate has yet to be discovered. Cross-linguistically, the most common (or neutral) realization of primary word stress involves a fall in pitch rather than a rise [11, p. 43], the phonetic basis for this being that a fall is articulatorily less complex than a rise. Furthermore, it seems to be a general characteristic of stress languages that in utterance-final position, stress is realized "as a high-low fall in pitch, rather than as a high level pitch" [11, p. 43]. In Germanic languages, however, duration and spectral energy distribution, not pitch, are regarded as likely candidates for reliable correlates of word stress [12]. Pitch is not considered an invariant correlate, since stressed syllables are not necessarily associated with pitch accent. For German, the experimentally established cues of word stress include vowel duration, pitch and intensity changes as well as laryngeal features [13 and 14]. In perception, however, it seems that German speakers use changes in F0 as the most powerful cues of prominence [15]. This finding supports Kager's claim [16, p. 367] that duration and intensity may be subordinated to pitch as "prominent syllables are potentially capable of bearing pitch movements with a strong perceptual load".

The psychological reality of stress violations in speech was found to correspond to specific neural and acoustic correlates in German. Knaus and colleagues [17] measured the brain activity of German subjects while the latter were presented with both correctly and incorrectly stressed (monomorphemic) words. The presentation of incorrect stress patterns was found to evoke a negativity effect (N400) reflecting an inhibition during the lexical retrieval process. Two ERP studies [18, 19] indicate that in the presence of a deviant stress pattern, German speakers' perceptual system is sensitive to foot structure. In particular, a stress shift inducing a total restructuring of the binary foot (e.g. *Lexíkon < Léxikon*) leads to stronger effects than shifts that do not change foot structure (e.g. *Lexikón < Léxikon*). Stress shift was also studied in several stress clash environments (i.e. after prefixation, composition and phrasing). Wagner and Fischenbeck [20] found that in compounds, stress shift (and stress reduction) is not only a perceptual phenomenon but also has an acoustic manifestation, namely in the F0 contours. When listeners do not perceive any stress shift, the presence of a late peak on the primary stress syllable is observed, whereas a flat F0 contour after the pitch accent correlates with the perception of stress reduction. There is thus some evidence that in speech, like in songs, perception of prominence relations in German is affected by both pitch movements and the parsing of syllables into constituents (feet in language, groups in song).

On the musicological side, we should carefully consider the nature of meter and accentuation in music. Lerdahl and Jackendoff [21] distinguish three types of accents that contribute to metrical surface structure in music: phenomenal, structural and metrical accents. In the present study, like in those that have preceded it, only metrical accents were encoded in the metrical grid and used to determine the relative

strength of the musical beats. However, the two other accents also contribute to structuring the beat pattern of a tune. Particularly important are phenomenal accents, which may confer emphasis or stress on a moment in the musical flow by means of long notes, leaps to relatively high or low notes, sudden changes in dynamics or timbre or local stresses such as *sforzandi*. Any of these events at the musical surface may function as a perceptual input to metrical accent, which may either contribute to or contradict the established sense of meter. Structural accents, i.e. accents "caused by the melodic/harmonic points of gravity in a phrase" [21, p. 17], occur at the beginning and the end of rhythmic groups (cadences). Cadences are special locations where the perceptions and expectations of tonal motion and resolution are heightened. The perception of metrical surface structure further relies on the underlying melodic structure (the so-called *pitch reduction*). For melody, Lerdahl and Jackendoff [21] posit a hierarchical, recursive structure very similar to the metrical structure found in rhythm. Like rhythms, melodies are structured in terms of different levels of abstraction. At an abstract level, a melody spells out a tonic chord; at another level, closer to the surface, it spells out a scale. In a melodic phrase, a note that belongs to the more abstract levels is understood as relatively stable, the points of highest stability usually being long notes. The shorter notes should instead be understood as transitions between stable, anchoring points in the phrase.

Comparing the musical structure of the tunes in Fig. 4, 5 and 6 provides useful insights into the relationship between the various hierarchical levels at work. In the original tune (Fig. 4, pos. 7 through 9), the transition from the dominant chord VII to the tonic chord I (at the abstract level, the so-called *resolution*) is realized through a major second interval (A4–G4), an interval that is particularly good for resolution in music, leading from a tensing to a relaxing motion. Changing this interval to B4–G4 right in the cadence in construct #1 strongly affects the perception of resolution. Moreover, the shape of the resulting melody is very different: while the original tune features a linear descent from B4 to G4, the melodic contour of construct #1 exhibits an ascending-descending pattern (A4–B4–G4), with the highest pitch in the whole phrase placed right in the cadence. In Thomassen's model of perceived melodic accent [22], more accent is generally associated with contour pivot pitches, i.e. pitches coinciding with changes of direction, and in particular with ascending-descending pivots. In construct #1, there is therefore a conflict between the metrical hierarchy, which assigns prominence to pos. 9, and the melodic accent, which inserts prominence earlier, at changes of direction (i.e. A4–B4 and B4–G4, pos. 7 through 9). In construct #2 (Fig. 6), a rhythmic change is introduced whereby the duration of the first A4 pitch is extended by one eighth note. As a result of this change, the stressed syllable *Weib.* is moved to a position (pos. 8) in which it surfaces not only as metrically weaker than in the original formula, but also as significantly shorter (i.e., eighth note < quarter note). This reduces its function as an anchoring point in the melody as well as its perceptual load, eventually turning it into an "ornamental" element less stable than the preceding and following syllables. When the text is grafted onto this rhythmical formula, well-formed in itself, a misalignment arises (and is clearly perceived) between linguistic prominence, constituent structure and meter.

The discussion of the settings in Fig. 4, 5 and 6 is meant as an exemplification of the intrinsic complexity of the internal rhythmical and melodic organization of music, which challenges the study of textsetting. In order to understand how language prosody interacts with musical rhythm in songs, it is in fact necessary to take into account the multiple levels of accentuation and emphasis available in the music. Moreover, in order to make valuable use of textsetting data to gain insight into the nature of linguistic rhythm, it is necessary to control for the various parameters (acoustic and perceptual) that affect the expression of prominence in singing. Although linguistic prominence in singing may be signaled by other cues than in speech, musical settings may be used as a magnifying glass showing how correlates of prominence interact with each other in a controlled environment. Ultimately, the unconscious knowledge that listeners access while making sense of sung texts could become a valuable source for our understanding not only of how textsetting works, but also of prosody itself.

4. Conclusions

This study provides partial support for the hypothesis that in German, a stress-based language, the SBM principle is strictly obeyed. It also shows, however, that the application of this principle alone cannot capture the complexity of textsetting in German (and possibly in other languages). In particular, it fails to make predictions about the setting of morphemes with ambiguous stress patterns (stress doublets), and to account for stress-beat mismatches lacking a perceptual correlate.

Moreover, manipulations of the melodic contour (construct #1) in correspondence to a mismatch location have indicated that native speakers' perception of stress-beat discrepancies relies not only on metrical factors, but also on melodic features, namely on differences in pitch interval. Manipulations of interval durations (construct #2) indicate that violations of constituent structure also impinge on the well-formedness of settings.

Overall, German textsetting data suggests further arguments in favor of a systematic study of prosody in vocal music.

5. Acknowledgements

6. References

[1] T. A. Morgan and R. D. Janda, "Musically-conditioned stress shift in Spanish revisited:empirical verification and nonlinear analysis," *Studies in Romance linguistics, selected proceedings from the XVII linguistic symposium on Romance languages* (ed. by C. Kirschner and J.A. Decesaris), pp. 273–288, 1989.

[2] J. Halle, Text, Tune and Metrical Form, chapter 3: What is a Mismatch. https://johnhalle.com/musical.writing.technical/comparative.text setting.pdf

[3] B. Hayes and A. Kaun, "The role of phonological phrasing in sung and chanted verse," *Linguistic Review* vol. 13, pp. 243–303, 1996.

[4] *Deusche Volkslieder. Texte und Melodien*. Stuttgart: Reclam, 2001.

[5] *Volkslieder. Die schoensten deutschen Volkslieder*. Stuttgart: Carus, 2010.

[6] F. Dell and J. Halle, "Comparing musical textsetting in French and in English songs," *Towards a typology of poetic forms* (ed. J.-L. Aroui and A. Arleo), pp. 63–78, 2009.

[7] J. Halle, *A Grammar of Improvised Textsetting*. Ph.D. dissertation, Columbia University, 1999.

[8] J. Halle and F. Lerdahl, "A generative textsetting model", *Current Musicology*, vol. 55, pp. 3–23, 1993.

[9] T. Proto and F. Dell, "The structure of metrical patterns in tunes and in literary verse. Evidence from discrepancies between musical and linguistic rhythm in Italian songs," *Probus*, vol. 25, pp. 1–34. 2013.

[10] R. Wiese, *The Phonology of German*, Oxford: Oxford University Press, 2000.

[11] L. M. Hyman, "On the nature of linguistic stress," *Studies in stress and accent. Southern California Occasional Papers in Linguistics 4* (ed. L.M. Hyman), pp. 37–82, 1997.

[12] A. Sluijter, *Phonetic Correlates of Stress and Accent*, HIL dissertations 15, Leiden University, 1995.

[13] K. Claßen, G. Dogil, M. Jessen, K. Masarek and W. Wokurek, "Stimmqualität und Wortbetonung im Deutschen," *Linguistische Berichte*, vol. 174, pp. 202–245, 1998.

[14] M. Jessen and K. Marasek, "Acoustic correlates of word stress and the tense/lax opposition in the vowel system of German," Phonetic AIMS, vol. 2, no. 2, pp. 143–146, 1995.

[15] K. J. Kohler, "The Perception of Prominence Patterns", Phonetica 65, pp. 257–269, 2008.

[16] R. Kager, "The Metrical Theory of Word Stress," *The Handbook of Phonological Theory* (ed. J. Goldsmith), pp. 367–402, 1996.

[17] J. Knaus, R. Wiese & U. Janßen, "The processing of word stress: EEG studies on task-related components", *Proceedings of the International Congress of Phonetic Sciences 2007, Saarbrücken*, pp. 709–712, 2007.

[18] U. Domahs, R. Wiese, I. Bornkessel-Schlesewsky, M. Schlesewsky, "Word prosodic hierarchies: evidence from event-related potentials," *Phonology*, vol. 25, pp. 1–36, 2008.

[19] J. Knaus and U. Domahs, "Experimental evidence for optimal and minimal metrical structure of German word prosody", *Lingua*, vol. 119, pp. 1396–1413, 2009.

[20] P. Wagner and E. Fischenbeck, "Stress perception and production in German Stress Clash Environments," *Proceedings of Speech Prosody 2002* Aix en Provence, France, 2002.

[21] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music*, Cambridge: MIT Press, 1983.

[22] J. Thomassen, "Melodic accent: Experiments and a tentative model," *Journal of the Acoustical Society of America*, vol. 71, pp. 1596–1605, 1982.