Chapter 1

Language and Music in Optimality Theory

1.1. Introduction[2]

Jackendoff and Lerdahl (1980) point out the resemblance between the ways both linguists and musicologists structure their research objects. This insight gave rise to the proposal of a formal generative theory of tonal music (Lerdahl and Jackendoff 1983), in which they describe musical intuition. Above all, insights from non-linear phonology (*cf.* Liberman 1975; Liberman and Prince 1977 among others) led to scores provided with tree structures, indicating heads and dependent constituents in the investigated domains. In this way, composer Lerdahl and linguist Jackendoff bring to life a synthesis of linguistic methodology and the insights of music theory. Gilbers (1987) shows that music theory in turn can be useful to describe linguistic rhythmic variability (*cf.* also Gilbers and Schreuder 2002). Further examples of musical and linguistic cross-pollination include among others Jacobson (1932), Guéron (1974), Liberman (1975), Attridge (1982), Oehrle (1989), Wallin (1991), Raffman (1994), Hayes and Kaun (1996), Hayes and MacEachern (1998), Patel (1998, 2003), Patel et al. (1996, 1997, 1998a,b), Repp (2000).

Liberman (1975) claims that in principle every form of temporally ordered behaviour is structured the same way (*cf.* also Gilbers 1992). If this claim is true, language and music should have much in common, since both disciplines are examples of temporally ordered behavior. In this chapter we offer additional arguments for this proposition. In both fields the research object is structured hierarchically and in each domain the important and less important constituents are defined. In Lerdahl and Jackendoff's music theory,

[2] This chapter is based on Gilbers and Schreuder (2002) which will also appear in two parts as Gilbers and Schreuder (in press) and Schreuder and Gilbers (in press). In Dutch it has appeared as Gilbers and Schreuder (2000).

these heads and dependents are defined by preference rules determining which outputs, i.e., the possible interpretations of a musical piece, are well-formed. Some outputs are more preferred than others. Preference rules, however, are not strict claims on outputs. It is possible for a preferred interpretation of a musical piece to violate a certain preference rule. This is only possible, however, if violation of that preference rule leads to the satisfaction of a more important preference rule.

This system of violable output-oriented preference rules in music theory has been very familiar to linguists since 1993, for a practically identical evaluation system, which uses similar well-formedness conditions, can be found in Prince and Smolensky's Optimality Theory (1993) (further OT). This theory, first introduced in phonology, owes a great deal to the work of Lerdahl and Jackendoff. Currently, it is a leading phonological theory and is expanding from phonology to other linguistic disciplines. In OT well-formedness conditions on outputs, constraints, also determine grammaticality. Here, too, the constraints are not strict, but soft, or violable. However, a crucial difference between Lerdahl and Jackendoff's violable constraints and OT's seems to be in the nature of the rule interactions. In Lerdahl and Jackendoff (1983), unlike standard OT, rules are not strictly ranked, because they apply with variable strength, and because sometimes several weaker rules can gang up on a stronger rule. The Lerdahl and Jackendoff theory is more like the theory of Harmonic Phonology, a predecessor of OT. Recent accounts of OT, however, have loosened the requirement of strict dominance. Through variants like constraint demotion (Tesar and Smolensky 1998) or the Gradual Learning Algorithm (Boersma and Hayes 2001), constraint rankings can vary to some extent (*cf.* Chapter 3). In this chapter we will show that in the present state of phonology the resemblances are even more striking than in the time of Lerdahl and Jackendoff (1983).

The remainder of this chapter is constructed as follows: section 1.2 of this introductory chapter is further devoted to the resemblances between Lerdahl and Jackendoff's music theory and OT, with subsections on structuring, conflicting preference rules, and boundary marking. Section 1.3 gives our conclusion in relation to the study of temporally ordered behaviour.

1.2. The resemblances between language and music

In their 'Generative Theory of Tonal Music' Lerdahl and Jackendoff (1983) describe how a listener (mostly unconsciously) constructs connections in the perceived sounds. The listener is capable of recognizing the construction of a piece of music by considering some notes/chords to be more prominent than others. This enables him for example to compare various improvisations on one theme and to relate them to the original theme, even if he does not know the original theme. It enables him to get to the bottom of the construction of a complete piece, as well as the constructions of the different parts of that piece. Where does a new part start? What is its relation to a preceding part? Which are the most prominent notes in a melody?

Our cognition thus works in a way comparable to how a reader divides a text (often unconsciously too) into different parts. A reader also distinguishes paragraphs, sentences and constituents. He structurally divides a text. What is the nucleus of a sentence? What is attributive and therefore less prominent?

The term 'language' as used in this dissertation has a very broad meaning. We mean any module of the language faculty which deals with hierarchical structure and which can be analyzed as consisting of deconstructable parts which stand in hierarchical relationships to each other, i.e. grammar. This contains at least syntax, morphology and phonology as it is represented in our unconscious knowledge.

In section 1.2.1 we will show what the resemblances are between language and music with regard to the division of the research object into smaller domains. Section 1.2.2 is about the resemblances in well-formedness rules, which are output-oriented, and which determine the main constituent and the dependent constituents for each domain.

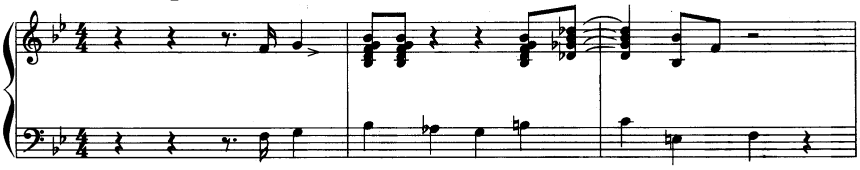
1.2.1. Structuring

In music theory the musical stream of sounds is hierarchically divided into structural domains. Each domain contains some smaller domains, which in turn contain smaller domains. The smallest domain in music is the motif (built up out of notes), a short, rhythmic, melodic or harmonic building block, which is a recurrent element in the whole piece of music. It retains its identity when elaborated on or transformed and combined with other material (Randel 1986: 513). Several motifs together form phrases, and phrases together may build up themes. A phrase, or period, is a kind of musical sentence, which concludes with a moment of relative tonal and/or rhythmic stability such as produced by a cadence, *cf.* section 1.2.2.4 (Randel 1986: 629). The realization of phrasing in performance is largely the function of the performer's articulation. A theme is a musical idea, usually a melody, that forms the basis for a composition of a major section of a composition. It can consist of a single phrase or several phrases together (Randel 1986: 844). It generally covers several measures and is regularly varied upon during the whole piece. In principle the listener is always able to recognize the theme, although it can be somewhat different each time. He reduces every occurrence of the theme to its underlying structure. The motifs and themes together determine the character of the piece of music. Several phrases or themes can form a section or verse, etc. By imposing this hierarchical structure on the entire piece, the listener is able to understand it. Figure 1 shows an example of the construction in the jazz original 'Tuxedo Junction'.

Figure 1 Tuxedo Junction

a. Motif

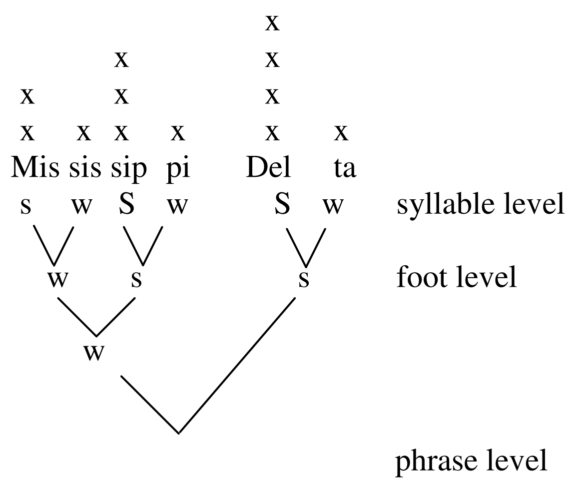
b. Theme or phrase



c. Section

Comparable domains can be found in language. The building block in language comparable to the motif in music is the morpheme (built up out of phonemes). Morphemes are joined together into larger meaning-bearing units: words, compounds, constituents (phrases), etc. And just as we have a rhythmic division (*metrical structure*) in addition to a melodic division (*grouping structure*) in music, we can divide rhythm in language into syllables as well, united into feet, which are comparable to the musical measure. In language – as in music – this division of the sound signal into domains allows us to grasp the structure and to understand how to interpret the whole text. Figure 2 shows an example of a structured phrase in language. The height of the grids reflects the degree of stress and the tree diagram represents the relative strength between the syllables and feet.

Figure 2 Prosodic construction of a phrase (Prince 1983)



1.2.2. Conflicting preference rules

1.2.2.1. Evaluation of possible output candidates

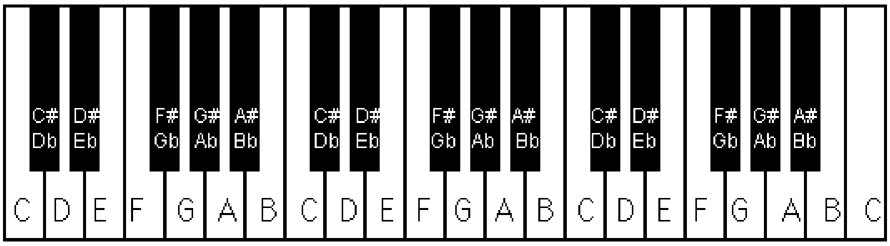
In language (Prince and Smolensky 1993) as well as in music (Lerdahl and Jackendoff 1983) the head of each domain is chosen by means of well-formedness conditions. A coherent whole of such conditions (or constraints) indicates what is grammatical in language and which mode of perception is optimal in music. In language for example one has to know which of two syllables in a foot is stressed and in music which chord of a certain sequence is the most prominent in the progression of the whole piece.

Possible candidates for every output form are evaluated by the constraints. These constraints can be contrasting and lay down opposite requirements on the output structure or interpretation to be preferred. Conflicts are thus solved by assuming differences in weight between the different constraints. In this way a weight hierarchy of constraints is arranged. One could compare this to traffic rules. Traffic coming from the right has priority, unless the traffic coming from the left is driving on a major road. This last rule, however, is overruled by the rule stating that one has to wait for a red traffic light. In traffic we are dealing with a collection of hierarchically ordered rules. Note that these rules are soft. They can only be violated in order to satisfy a higher preferred rule (minimal violability).

Linguistic constraints in OT are soft too. An output candidate can be grammatical, even if it violates constraints. As long as no better candidate comes up, the least bad candidate is the optimal one. Suppose we have a word with two syllables CVCVVC (*papaap*) and we have to determine on which syllable stress falls, given two relevant constraints: a positional constraint i (stress never falls on the last syllable) and one in which syllable weight plays a role, constraint j (stress falls on the heaviest syllable). The best output according to constraint i is then: *pápaap*, but *papáap* is the best according to constraint j. There is no output which satisfies both constraints. In a grammar conflicts like these are solved by a language-specific ranking of the constraints according to their importance. These universal constraints are not ranked in themselves, but in the grammar of a particular language they are strictly ordered. A language learner has to acquire the knowledge that in language X constraint i has priority over constraint j, while in language Y it can be the other way around.

The well-formedness rules in music theory are also potentially conflicting and soft. One of the conditions implies that a chord in a metrically strong position (for example the first beat in a measure) is more important than a chord that is not in such a position. A chord in a metrically strong position is preferred by the listener to act as most prominent chord (the head) of the measure or the phrase, above all other chords in the same sequence. Another preference rule states that, given the tonality of the piece, all chords are harmonically unequal in their strength. In a piece in the key of C, the G-chord is harmonically more consonant than a B-chord. Thus there will be a conflict between preference rules if a B-chord is in the first position of a measure and a G-chord is in the last. Lerdahl and Jackendoff solve this kind of conflict by hierarchically ranking the preference rules. In our example the preference of a harmonically more consonant chord outweighs the preference of a metrically stronger chord, so that the listener will choose the G-chord as head and not the B-chord, given the key C.

Figure 3 Piano keyboard



An apparent difference between music and language is that Lerdahl and Jackendoff give only one ranking of well-formedness rules, while in OT a ranking of the universal constraints, in themselves unranked, has to be made for every language. Although Lerdahl and Jackendoff only offer one ranking for tonal music, one can imagine that, for example, prolongation of a melodic line is relatively more important in Eastern music than in Western music, while possibly in Western music relatively more weight is attributed to harmonic consonance of a piece. Perhaps differences in musical styles can be accounted for in the same way as for differences between languages (*cf.* also Patel and Daniele 2003, and Chapter 2).

In the next subsections we will discuss two examples of a conflict between positional and segmental markedness. In section 1.2.2.2 we present a linguistic example based on language acquisition data; in section 1.2.2.3 a comparable example in music is given.

1.2.2.2. A linguistic example of conflicting constraints: language acquisition

The language acquisition data in Table 1 prove that several kinds of markedness play a role in the acquisition of clusters. In this example we can see a conflict between segmental markedness and positional markedness in the realizations by the Dutch boy Steven of respectively *acht* 'eight' and *korst* 'crust'.

Table 1 Cluster reduction Steven

age:	target word:	input:	realisation:
1;11	*acht*	/ɑxt/	[ɑt]
2;2	*korst*	/kɔrst/	[kɔs]

Data: Van der Linde (2001)

The dominating constraint in both cases is *COMPLEX, a prohibition on consonant clusters in the output. Prince and Smolensky (1993) propose HMARG to indicate that in marginal syllable positions less sonorant segments are prefererred to more sonorant ones. The child has arrived at a phase of its development in which the correspondence constraint MAX I-O, a constraint which demands that every segment of the input has a correspondent in the output, and therefore forbids deletion, is dominated by *COMPLEX and HMARG. With the help of these constraints we get to the analysis in Table 2.

Table 2 Provisional OT analysis

constraints → /ɑxt/ candidates ↓	*COMPLEX	HMARG	MAX I-O
[ɑχt]	*!		
[ɑχ]		/χ/!	*
☞ [ɑt]		/t/	*

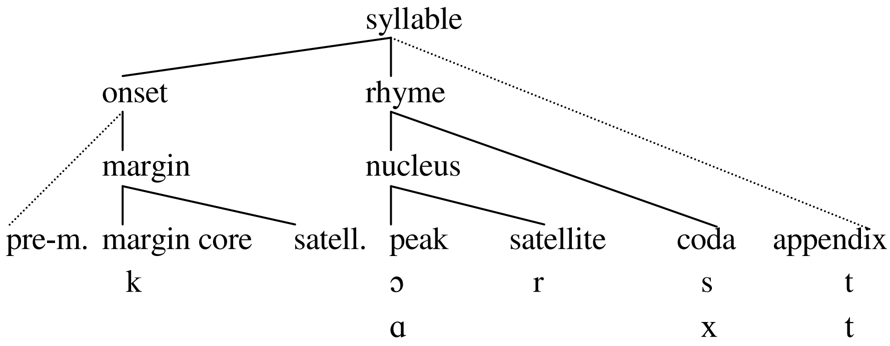
The constraint ranking in Table 2, however, wrongly predicts that the realisation of *korst* would be [kɔt]. We assume that Steven's realisation [kɔs] should be explained by the supposition that the difference between the syllable positions of /t/ and /s/ has its influence. HMARG is violated to satisfy a higher-ranked constraint with respect to positional markedness.

A straightforward CVC-syllable model and constraints like *COMPLEX and *CODA (syllables must end in a vowel) are not satisfactory for describing phonotactic restrictions and positional markedness relationships between segments in a Dutch syllable. We therefore copy a more complex syllable template in Figure 4 from Gilbers (1992). This model is based on a proposal in Cairns and Feinstein (1982), in which differences in positional markedness are stipulated, mixed with a proposal in Van Zonneveld (1988), in which an X-bar theory for syllable structure is developed[3].

---

[3] Cairns and Feinstein indicate differences in markedness between consonant sequences like obstruent–liquid; obstruent–nasal. Unfortunately their model lacks sequences with fricatives such as in *schaap* [sχa:p] 'sheep'.

Figure 4 Syllable template



The model in Figure 4 represents a hierarchical organization of the segmental distribution in a syllable. The vertical lines indicate the head of a branching constituent and the slanting lines the dependent parts. Thus the Margin core is the head of the Onset, which is dependent on the Rhyme constituent. In OT we express this hierarchical structure in a series of ranked universal constraints. A Satellite always takes a more marked position than a Coda. The most marked positions are the Pre-margin and the Appendix, the so-called extra-syllabic positions (X-SYLLABICITY). Table 3 represents the ranking of the relevant constraints. The order is universal, but other constraints can be placed in between the various positional constraints.

Table 3 Positional markedness

*X-SYLLABICITY >> *SATELLITE >> *CODA

In the original model subcategorization rules were given for the contents of the various syllable positions. Thus in the nucleus position only vowels can occur and in the satellite positions only sonorant consonants are allowed. In OT, however, all constraints are violable and we therefore state that a SATELLITE prefers sonorant consonants above other consonants. All positions in the syllable template can in this way be formulated as OT constraints, using their subcategorization preferences as the violable rule. In an optimal parsing of *acht*, /x/ takes the coda position and /t/ the appendix position.

Steven's realisations can be described by means of the tableaus in Table 4, based on the Constraint Demotion Algorithm for language acquisition by Tesar and Smolensky (1998). In Table 4a we see that before his second birthday Steven is in a phase in which segmental markedness (HMARG) dominates positional markedness (*XSYLL, *SAT), but that after his birthday positional markedness has become more important than segmental markedness. Finally, the correspondence constraints will dominate the markedness constraints. Phonological development is then completed.

Table 4 Analysis *acht* and *korst*

a. table for *acht* (phase Steven (1;11))

constraints → /ɑxt/ candidates ↓	*COMPL	HMARG	*XSYLL	*SAT	MAX I-O	*CODA
[ɑχt]	*!	/χt/	*			*
[ɑχ]		/χ/!			*	*
☞ [ɑt]		/t/	*		*	

b. table for *korst* (phase Steven (2;1))

Constraints → /kɔrst/ candidates ↓	*COMPL	*XSYLL	*SAT	HMARG	MAX I-O	*CODA
[kɔrst]	*!	*		/rst/		*
[kɔr]			*!	/r/	**	
☞ [kɔs]				/s/	**	*
[kɔt]		*!		/t/	**	
[kɔrs]	*!		*	/rs/	*	*
[kɔrt]	*!	*	*	/rt/	*	
[kɔst]	*!	*		/st/	*	*

Notice that OT is not a theory on representations or models. Table 4 is based on the model in Figure 4, where /t/ is not a Coda because it is in an Appendix position, and /r/ is not a Coda because it is in the Satellite position.

In music conflicts also arise between positional and 'segmental' markedness. In the next subsection we give an OT analysis of a passage from Mozart.

1.2.2.3. A musical example of conflicting constraints: OT analysis of Mozart K. 331, I

In music, similar to language, different preference rules can be arranged, in order to solve conflicts concerning the head of a domain. Segmental markedness has its musical equivalent in the hierarchical relationships between notes in a given tonality. Positional markedness is comparable to the strength differences between different positions in a measure.

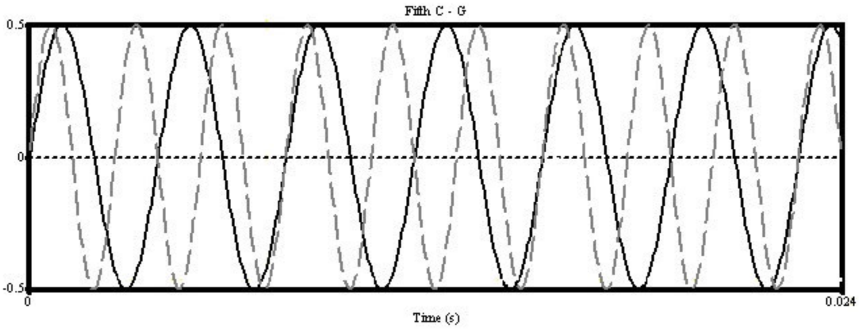
With regard to segmental markedness, musical segments – like segments in language – keep hierarchical relationships with each other. The hierarchy of musical segments, the pitches, is connected to the tonality of the piece. In tonal music, every piece is based on a given scale (the key or tonality of the piece), which means that all notes are arranged around the most important notes in that scale; the melody usually ends in the tonic, the keynote of that scale.

The tones of the scale can be combined in several ways, following each other in a melody, or harmonizing in chords. One harmony or succession sounds 'better' than the other. Intervals that are stable and do not require resolution are called 'consonant', more complex sounding intervals are called 'dissonant'. Dissonant harmonies are regarded as having an instability that requires resolution to a consonance (Randel 1986: 197). Like sonority in language, consonance and dissonance are gradual concepts. The hierarchical division of pitches in a piece happens on the basis of the relative consonance (Lerdahl and Jackendoff 1977, 1983). A relative consonant tone in the key of the piece is higher in the hierarchy than a relatively dissonant tone.

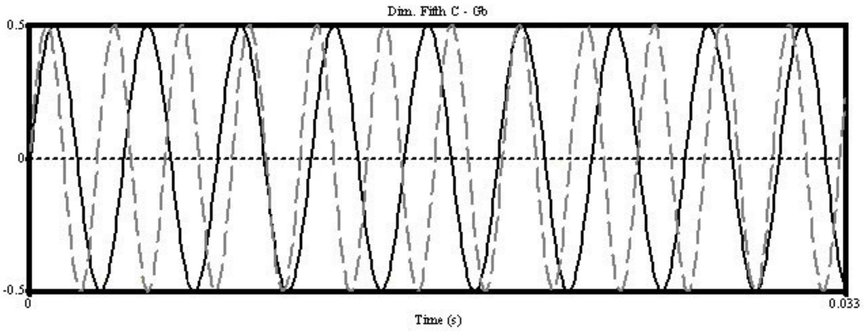
That consonance and dissonance are not a matter of taste, but a matter of acoustics, is shown in Figure 5. Consonant intervals consist of a simple ratio, whereas dissonant intervals have a more complex ratio. The ratio in e.g. a fifth is 2:3, as illustrated by two wave forms in Figure 5a. The three cycles of the G cross the C sinus in the zero boundary after two cycles of the C. The wave sinuses in Figure 5b, on the other hand, are much more complex; the sinus crossings do not intersect with the zero line anywhere. The extent of complexity of the wave ratios corresponds to the perception of relative consonance.

Figure 5 Consonance and dissonance

a. Consonant: perfect fifth C-G



b. Dissonant: diminished fifth C-G*b*



In addition to segmental markedness, there is also positional markedness in music, just as we saw in Figure 4. The first position in a measure is stronger than the second, and in for example the 4/4-measure the third position is less strong than the first, but stronger than the second or the fourth.

Lerdahl and Jackendoff developed the so-called time-span reduction, a kind of tree and grid construction, based upon the metrical structure and the grouping structure of a part of a musical composition, so as to reflect the hierarchical relationships between all pitches in relation to the tonality of the piece (see Figure 7). These relationships are determined by application of the preference rules, which determine the head in each domain. The head of a time-span Z is selected from the heads of the time-spans directly dominated by this time-span Z. The subordination relationship is transitive here: if X is an elaboration of Y and Y of Z, than X is also an elaboration of Z. Lerdahl and Jackendoff (1983) treat nine time-span reduction preference rules (TSRPR). Table 5 gives three examples of such rules.

Table 5 Time-span reduction preference rules

TSRPR 1:	Choose as the head of a time-span the chord (or the note) which is in a relatively strong metrical position (positional markedness).
TSRPR 2:	Choose as the head of a time-span the chord (or the note) which is relatively harmonically consonant (segmental markedness).
TSRPR 7:	Choose as the head of the time-span the chord (or the note) which emphasizes the end of a group as a cadence (comparable to the boundary marking effect of alignment constraints in language, *cf.* Table 7).

An example of a strong metrical position from TSRPR 1 is the first position in the measure. TSRPR 2 is connected to a hierarchy of chords based on harmonic stability. A triad tonic–tierce–fifth (c-e-g) is more stable than a seventh chord (c-e-g-b flat), while a seventh chord in its turn is more stable than for example a suspended fourth (sus4) (c-f-g). The optimal chord according to TSRPR 7 is the final chord, a chord which generally is built on the tonic, preceded by a dominant chord (see Figure 11a). In C the dominant is G. Each smaller group concludes with a chord suitable for a cadence. There are also ‘lighter’ cadences, however, indicating that a group is not definitely concluded, and that the melody will continue after the cadence, moving to a next group. Often the sequence subdominant–tonic is used (the plagal cadence, F-C, *cf*. Figure 11c). The first three positions in the harmonic hierarchy are occupied by the tonic, the dominant, and the subdominant respectively.

As in OT the set of preference rules from music theory is hierarchical. TSRPR 2 is stronger than TSRPR 1; TSRPR 7 is stronger than TSRPR 1 and TSRPR 2 together. In Figure 6 we give the first movement from a sonata by Mozart.

Figure 6 Mozart: Sonata K. 331, I (Lerdahl and Jackendoff 1977)

For this part we can determine the heads by means of application of the TSRPR hierarchy. The first four measures from the piece form the first group. In measure 3 the A^6-chord (F#-E-A) is metrically the strongest chord, and thus the head. In measure 4 the E-chord (E-G#-B) is the head, because it marks the end of the whole first group of four measures. Now the head has to be chosen for the group which is formed by measures 3 and 4 together. Metrically speaking, the A^6-chord is still the strongest. But TSRPR 7 dominates TSRPR 1. In Table 6 we give an example of an OT-like musical analysis. Although the A^6-chord is metrically speaking in a stronger position than the E, the dominant TSRPR 7 prefers the dominant chord E as the cadence in this phrase.

Table 6 OT analysis

constraints → A^6 – E Candidates ↓	TSRPR 7	TSRPR 2	TSRPR 1
☞ E			*
A^6	*!	*	

This choice has consequences for the tree in Figure 7, in which the E-chord dominates the A^6-chord. The E-chord in its turn is dominated by the harmonically more consonant initial A-chord of the piece, and at the top of the hierarchy is the final chord of the whole group of eight measures, again an A-chord, because it is the head according to both TSRPR 1 and TSRPR 7.

Replacing all notes/chords which are chosen as heads of every time-span by gridmarks shows the resemblance to metrical phonological representations as proposed in Liberman (1975), Liberman and Prince (1977) and Hayes (1984) among others (see Figure 8). The underlined gridmarks ($\underline{x}$) indicate *silent beats* (*cf*. Selkirk 1984). Silent beats are filled either by a rest or by lengthening of a preceding note. Note that metrical grids usually indicate stress differences, whereas this grid indicates prominence differences between chords, not stress. Obviously, the same kind of representations can be used to indicate differences in prominence.

The analysis shows that the beginning and end of the phrase are emphasized. TSRPR 7 dominates the constraints referring to segmental and positional markedness. In language, boundaries of a phrase may also be emphasized. In this way a stress shift as in *Mississìppi Délta*, realized in fast speech as *Mìssissippi Délta* (Hayes 1984), can be described (*cf*. Visch 1989). We will examine this in the next subsection. For an elaborate experiment with regard to boundary alignment we refer to Chapter 4.

Figure 7 Time-span reduction (Lerdahl and Jackendoff 1977)

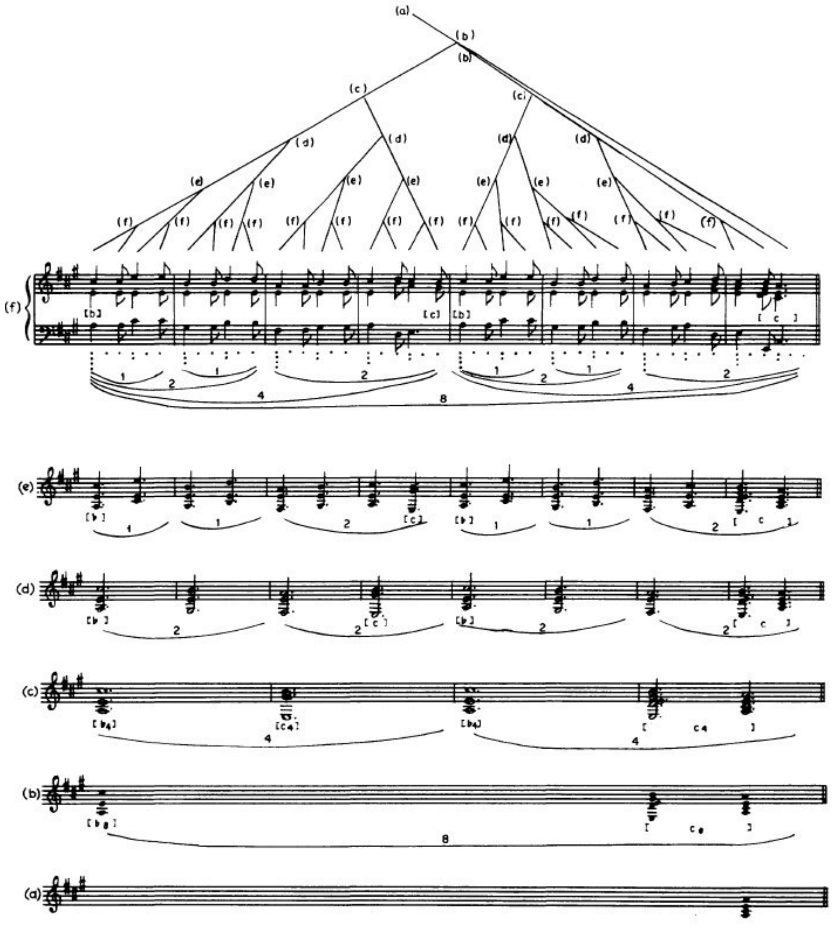


Figure 8 Grid of the time-span reduction in Figure 7
(two adjacent phrases)

```
                                             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
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
```

1.2.2.4. Boundary marking

In both music and language, several processes can be considered to be boundary markers. Secondary stress shift (or 'early accent placement' as we call it in Chapter 4) and final lengthening are such processes. In OT so-called generalized alignment constraints are proposed for the analysis of boundary marking processes (McCarthy and Prince 1993b). All alignment constraints refer to constituent boundaries, and they have the following form:

Table 7 *Alignment*

Align (Cat 1, Edge 1, Cat 2, Edge 2) =
∀ Cat 1 ∃ Cat 2 where Edge 1 of Cat 1 and Edge 2 of Cat 2 coincide

Alignment constraints prefer output candidates in which for example a constituent boundary coincides with a stressed syllable or in which a morphological boundary coincides with a phonological one.

A predecessor of alignment constraints for the controlling of rhythmical boundary marking in language is the Phrasal Rule of Hayes (1984). Hayes gives examples of preference rules for an ideal rhythmic structure in language: Eurhythmy rules. He attributes rhythmic shift to adjustments to ideal patterns for rhythmic sequences. The Phrasal Rule (PR) is one of these Eurhythmy rules. It implies that a grid is more eurhythmic if it contains two marks as far apart from each other as possible, at the second-highest level. The PR makes that the boundaries of the phrase are emphasized. Van Zonneveld (1983) called this phenomenon 'Rhythmic Hammock'.

Table 8 Rhythmic Hammock in *individualistisch persoon* 'individualistic person' (Visch 1989: 102)

constraints → *individualistisch persoon* candidates ↓	HAMMOCK	CORR
individualìstisch persóon	*!	
☞ *ìndividualistisch persóon*		*

In Table 8 the hammock pattern is visible the second candidate. This pattern is comparable to the grid pattern in Figure 8 for the music passage in Figure 7, where the extremes are marked by the highest grid columns. Because Hammock, like TSRPR 7 in music, is a dominant constraint, the second candidate in Table 8 wins. So like similarities in segmental and positional markedness we also see a great similarity between language and music in the way boundaries are marked. Hammock patterns are found in phonology as well.

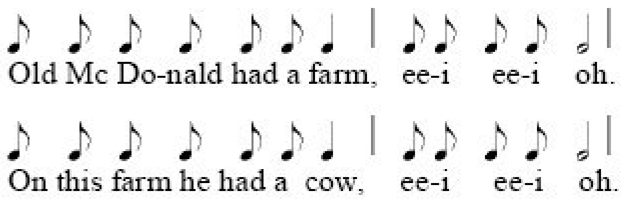
Another form of boundary marking which we find both in language and music is Final Lengthening (FL) (Lindblom 1978, Ladd 1996). FL is the phenomenon of lengthening of a note or a speech sound at the end of a phrase. According to experimental research by Lindblom (1978) in spoken Swedish the duration of the vowel [ɑ:] in ['dɑ:g] is longer at the end of a phrase (Table 9a) than when the word is in another position (Table 9b).

Table 9 Final Lengthening in Swedish

a. *finurlige Dag*	'ingenious Dag'
b. *Dag berättar*	'Dag tells a story'

In Table 9a the vowel is in final position and therefore it lasts ± 55 ms longer than in initial position, as in Table 9b. Figure 9 shows an example of FL in music, where it is a very common phenomenon.

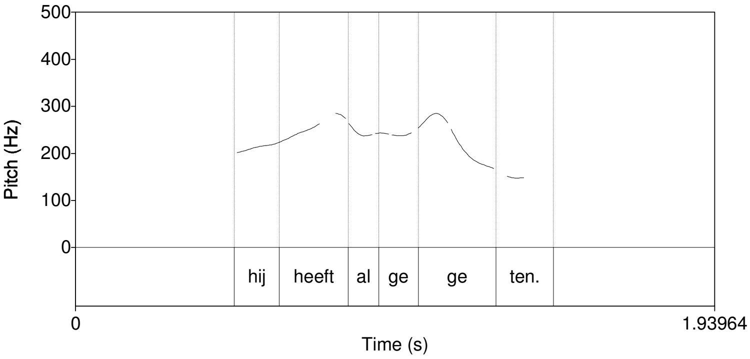
Figure 9 Final Lengthening in music (after Liberman 1975)



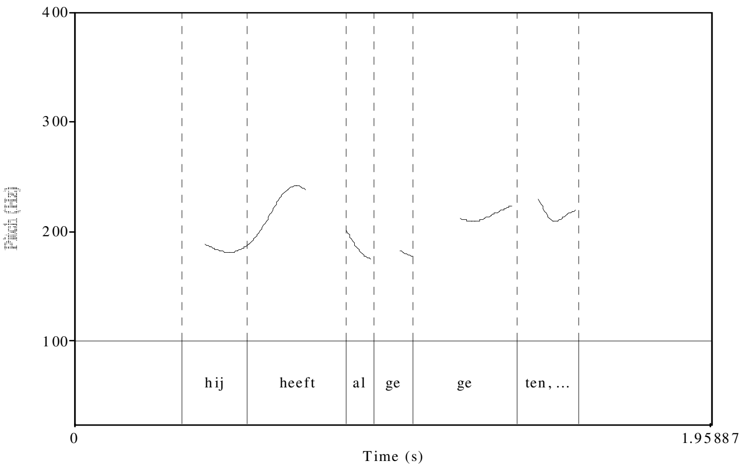
The last note of each phrase is lengthened indicating that the phrase is concluded. Another common phenomenon for marking boundaries in both language and music is deceleration of tempo (ritards) towards the end of phrases, as well as acceleration at the beginning of each melodic movement (Repp 1990, 1998). Patterns of tempo modulations often indicate a hierarchy of phrases, with the amount of slowing or phrase-final lengthening at a boundary reflecting the depth of embedding (Todd 1985, 1989, Palmer 1989, 1997, Repp 1990, 1998, 2002, H. den Ouden 2004). One can see that this gradation in FL occurs in Figure 9, as the note 'before the comma' is lengthened compared to the preceding notes, but less than the final note of the phrase.

In addition to rhythmic phenomena, intonation patterns are used to mark boundaries. In language, intonation marks groups such as syntactic constituents and phonological phrases. In a similar way intonation marks, for example, the differences indicated in writing by full stops and commas. A full stop in a declarative sentence is often the equivalent of a strong pitch fall in prosody, while a comma is comparable to the intonation pattern in which the tone is suspended somewere 'in between', to indicate that the sentence is to be continued (Swerts 1994, Van Donzel 1997, 1999). The contours in Figure 10 reproduce this difference.

Figure 10 Intonation patterns (Schreuder 1999)



a. *Hij heeft al gegeten.* ‘He has already eaten.’



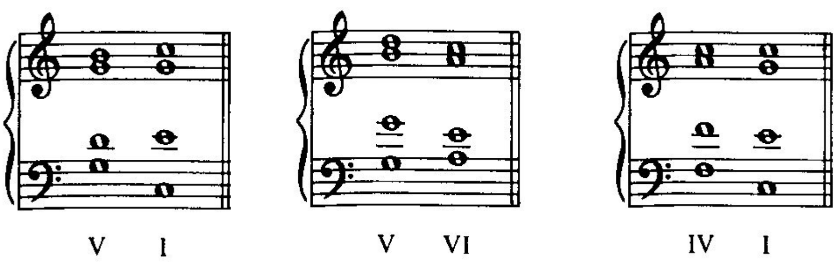
b. *Hij heeft al gegeten, (maar hij wil toch nog een koekje.)*
‘He has already eaten, (but he still wants another cookie.)’

In Figure 10a the intonation contour moves downward towards the end, creating a ‘final fall’, and in Figure 10b, the ‘comma intonation’, the tone is suspended in between (it is rather high in this example), the ‘continuation rise’. Thus in Figure 10b the sentence cannot be complete, something has to follow this boundary. The boundary tone of a question often rises to the top of the speaker’s range, although in the case of question intonation non-finality may not be the reason for the rise of tone (Gussenhoven 2002, 2004).

Intonation in language equals phrasing in music.[4] It causes the music to 'tell a story', similar to the way intonation does in language. Phrases are formed in which tension is built up or reduced, ending in a cadence, a melodic or harmonic configuration that creates a sense of repose or resolution (Randel 1986: 120). This is properly comparable to 'comma intonation' and 'full stop intonation' in language: the comma indicates prolongation, the full stop completion.[5] A full stop is comparable to the 'full cadence' (the end of a phrase or piece) in music, i.e. the sequence of G-C in the key of C, as in Figure 11a. Phrases and pieces prefer ending in the tonic, here C, mostly low.

Figure 11 Cadences

a. Full cadence b. Deceptive cadence c. Plagal cadence



A comma is comparable to a chord change in which the phrase does not end in the tonic, but e.g. in the fourth step of the scale (a 'deceptive cadence', as in Figure 11b), F in the key of C, so higher. It therefore does not sound completed, and another phrase, resolving in the tonic, will ideally follow. While in speech intonation the non-final boundary tone is mostly higher than the final boundary tone, in music this is no more than a tendency, because both boundary tones

---

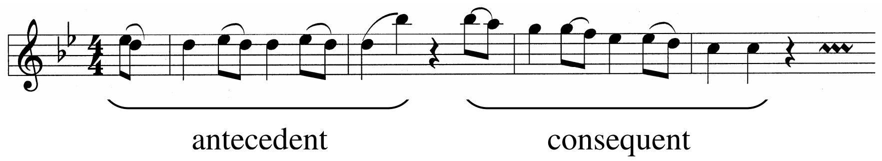
4 The term 'intonation' in music is reserved for tuning. We do not use the meaning 'tuning' in this dissertation.

5 Lerdahl and Jackendoff (1983) describe the difference between intonation patterns expressing prolongation and intonation patterns expressing completion. Prolongation is worked out in the prolongation reduction of the pitch structure.

can be the same tone. Depending on the melodic lines, and the harmonic progression of different cadences, the suggestion of a comma can nonetheless be evoked, as exemplified in e.g. the Marseillaise.

In Figure 12 we show a musical example of phrasing: ‘question and answer’, or ‘antecedent and consequent’. One phrase (the answer or consequent) follows the other phrase (the question or antecedent) and is also a reaction to it. The two phrases often have the same or similar rhythms, but have complementary pitch contours, e.g., a rising contour in the first and a falling contour in the second. An example is given in the first three measures of Mozart’s 40th Symphony in G Minor, K. 550:

Figure 12 Mozart K.550 (fragment): antecedent and consequent



This ‘question-answer intonation’ has a way of indicating grouping boundaries that is parallel to the patterns of full stops and commas. Again it is very similar to the patterns appearing in language. Questions have the tendency to end ‘upward’, while answers, comparable to sentences with full stop, tend to show a strong final fall. In fact, the example in Figure 12 shows this relationship at two levels simultaneously. This antecedent-consequent pair is followed by a similar pair one tone lower. At the same time, the two pairs are also related to one another as antecedent and consequent (Randel 1986: 42).

In this section, we showed that language and music have many similarities both on a representational level and in the sphere of preference rules. It seems that output-oriented preference rules do not specifically hold for only one discipline. In Chapters 3, 4 and 5, we will see that insights from music theory can be very useful in phonological issues.

With this in mind, we want to add a remark about the division into temporally organized elements, such as segments, accents, rhythm, or chords, and holistic patterns like intonation contours of phrases and melodies, all of which has to do with cerebral hemisphere specialization. In general, language processing is known to be situated in the left hemisphere. Only intonation is one of the few properties of language that are processed in the right hemisphere. Platel et al. (1997) and Stowe et al. (2005) point out that music perception is also located in both hemispheres; temporal patterns, like rhythm and chords, are located in the left hemisphere, and sequences of tones, i.e. melodies, and timbre, in the right. So intonation is literally the melody of language. The left hemisphere seems to be specialized in linear processes and consequently in analyzing temporal structures, of which rhythm, accents, segmental and positional hierarchies, etc. are examples. It was found that rhythm in music is processed by Broca's area, one of the neurological areas which are specialized in language processing. The right hemisphere, on the other hand, analyses in a holistic manner. It processes complex relationships and perceives patterns as units, instead of as sums of individual parts. These findings are highly controversial, however.

In spite of the controversy on this subject, it should be noticed that these alleged differences in processing of the two hemispheres reflect the differences between Lerdahl and Jackendoff's time-span reduction and the prolongational reduction. The prolongational reduction also analyses parts of a melody as larger units, not in a bottom-up fashion like the time-span reduction, but top-down. This might be new evidence for Lerdahl and Jackendoff's separation of the time-span reduction and the prolongational reduction (not elaborated in this dissertation. This separation has psychological relevance. It also shows that their theory, and especially OT, may give a good model for the way our brains work. We should keep in mind that OT has its source in connectionism. Connectionist models were developed in an attempt to construct a model that closely resembles the structure of the human brain (*cf.* Gilbers and De Hoop 1998).

1.3. Summary and Conclusion

In this chapter we followed Lerdahl and Jackendoff (1983), who observed a resemblance in the way musicologists and linguists structure their research objects. This observation led to their book 'A Generative Theory of Tonal Music', in which they describe music by means of a linguistic methodology: they used trees and grids, very common tools in syntax and phonology. Using these representations, they could visualize the hierarchical structures of which music is built up.

The methodology of preference rules Lerdahl and Jackendoff introduced to describe the way of achieving the ideal interpretation of a musical piece was followed ten years later by Prince and Smolensky (1993) in their Optimality Theory. The violable OT constraints show a striking similarity to the preference rules for music in defining the optimal output or musical interpretation. In this chapter we pointed to this similarity and we compared some of the musical preference rules to OT constraints.

We showed in this chapter that language and music also have much in common with respect to psychological assumptions and structural properties. In both disciplines the 'grammar' imposes hierarchical structures on the sound signal. In both language and music, preference rules for ideal outputs indicate the head constituent and the dependent constituents of every part of the hierarchical structure. Together the preference rules or constraints indicate what is grammatical in language and which way of listening is optimal in music. Moreover, in both theories the preference rules are soft and potentially conflicting, which gives the theories their descriptive power.

We gave a musical and a linguistic example of a conflict between positional markedness constraints and segmental markedness constraints. In music this conflict must be solved in order to decide which chord is the most prominent and will survive in the reduction. In language the outcome of the conflict is crucial to which segment is most prominent and will survive the simple system of a child acquiring language.

The domains in musical and linguistic structures are analogous. Both are deconstructible into smaller building units. The boundaries of these domains are the areas of several processes, many of which are involved in boundary marking mechanisms. Common phenomena of boundary marking in both language and music, such as phrasal accents that shift to the left boundary of a phonological phrase, and also final lengthening at the right boundary, are regulated by alignment constraints. In language as well as in music, the initial element is important, at all levels. Final elements are important as well, which is illustrated by resemblance of cadences in music to the final fall or rise in linguistic prosody. This should be seen as the result of satisfaction of so-called generalized alignment constraints. With respect to rhythm, similar restructuring processes occur, which seem to be the result of constraints referring to the Obligatory Contour Principle (OCP), a prohibition on adjacency of identical elements (McCarthy 1986). These constraints take a prominent position in the constraint ranking.

In our view, the observation that language and music show so many similarities strengthens the hypothesis that the same structures and principles hold for all temporally ordered behavior (*cf.* Liberman 1975; Gilbers 1992). In addition we can refer to research by Lasher (1978), who describes patterns in ballet in a similar way to our description of language and music in this chapter. In her research of dancing patterns the main movements are also distinguished from dependent movements, for every part of the hierarchically structured research object. It is the way in which our brain works: our cognitive system structures the world surrounding us in a particular way in order to understand everything in the best way.

On the basis of these resemblances we will show that insights of music theory can help out in phonological issues. Three of such issues are subjects of the experiments in this dissertation: variable rhythm, variable phrasing structure, and emotional intonation. In the remaining chapters of this dissertation we report on these experiments.