Optimal groups in music

Petra Hendriks and Sybrand van der Werf

Abstract

In this article we develop a musical parser which is based on Lerdahl and Jackendoff's Generative Theory of Tonal Music (1983). Their preference rules are adapted to incorporate recent insights from linguistic theory, in particular from Optimality Theory. Crucially, Optimality Theory provides us with a conflict resolution mechanism which allows us to determine unambiguously which rule is to be obeyed in case of a conflict between rules. The musical parser takes the form of a computational model which determines an optimal grouping structure for any given sequence of pitch events. Its viability is tested empirically by comparing its results with the results obtained from ten human subjects.

1. Introduction

As is widely accepted, the human cognitive system tends to organize perceptual information into hierarchical representations. This tendency can be observed in cognitive domains as widely varying as language, music and vision. An important question is whether a common system underlies the perceptual organization in all of these domains. Our research focuses on similarities and differences between language and music. We aim to investigate whether there is a general coherent model underlying grouping in language and music.

Since the rise of generative linguistics, linguists have stressed the view that natural language must be organized in a hierarchical fashion. Sentences are made up of phrases, which can again be made up of smaller and smaller phrases. This hierarchical organization of language can be conveniently represented by tree diagrams. An important insight in musicology was the realization that music might also be organized in such a tree-like fashion (Lerdahl & Jackendoff, 1983). In their formal Generative Theory of Tonal Music (extended in Lerdahl (2001)), composer Lerdahl and linguist Jackendoff propose a system of rules which determine how listeners intuitively organize a musical piece. Many of their rules take the form of preferences. Among other things, these preferences concern the way people perceive certain notes as belonging together. Because Lerdahl and Jackendoff assume sequences of notes to be organized into groups, and groups to be organized into larger groups, their system of rules results in the hierarchical organization of a musical piece.

Although Lerdahl and Jackendoff's approach borrows heavily from generative linguistics, their preference rules are at odds with the inviolable rules of grammar that are assumed to determine linguistic structure. However, a recent development in linguistics is the proposition that linguistic structures can be described and explained best by means of a system of violable rather than inviolable rules. This view is embodied in Optimality Theory (henceforth OT), which was introduced into linguistics by phonologist Prince and mathematical physicist Smolensky (Prince & Smolensky, 1993/2004). According to OT, a grammar consists of a set of constraints on possible output forms. A well-formed structure (for example, a grammatical sentence or a possible word) is the optimal structure for a given input. The optimal structure is the structure that satisfies the total set of constraints best. Although linguistic constraints typically impose conflicting demands on the output, these conflicts can be resolved because the constraints differ in importance, or strength. When it is impossible to satisfy all constraints, it is more important to satisfy the stronger constraint than the weaker one. An optimal output need not satisfy all constraints perfectly, but only needs to satisfy these constraints better than other candidate outputs. As a result, OT constraints are violable, and express linguistic tendencies rather than unviolable principles.

2. Optimal groups in language

As an illustration of the interaction among constraints in Optimality Theory, let us look at the way syllables are pronounced. Among the constraints determining the pronunciation of syllables are the constraints NoCoda and Parse. NoCoda expresses the cross-linguistic tendency for syllables to end on a vowel. This constraint is violated by any syllable ending with a consonant (a syllable-closing consonant is called a coda). The constraint Parse asserts that every speech segment in the input must appear in the output. This constraint penalizes deletion of segments. Crucially, if a syllable ends with a consonant, the constraints NoCoda and Parse are in confict because one way to avoid an output ending with a consonant is to delete this consonant.

Suppose our mental lexicon contains the underlying (in English nonsense) form /batak/. Our grammar, in particular our set of constraints and their relative ranking, determines how this underlying form is structured and pronounced. The output can vary in several ways: by different placement of syllable boundaries (indicated by a dot), by deleting or inserting segments, etcetera. Among the possible outputs we find [ba.tak], [ba.ta], [bat], [ba], [b] and even silence. These candidate outputs are evaluated with respect to the ranked set of constraints of our language, in our somewhat simplified example by NoCoda and Parse. Constraint evaluation in OT is usually displayed in graphic form by a constraint tableau (figure 1).

Input: /batak/	NoCoda	Parse
ba.tak	*!	
☞ ba.ta		*
bat	*!	**
ba		**!*

Figure 1. A constraint tableau in Optimality Theory.

The input and a few of the candidate outputs are listed in the first column. The constraints are listed in order of descending strength from left to right across the first row. An asterisk indicates a constraint violation. The exclamation mark indicates a fatal violation: a violation that renders this candidate suboptimal.

If NoCoda is stronger than Parse, it is more important to satisfy NoCoda than to satisfy Parse. The candidate output [ba.tak] violates NoCoda because the second syllable ends with the consonant [k]. For similar reasons, the candidate output [bat] violates this constraint. Violation of NoCoda renders these two candidates suboptimal. There are other candidates that do not violate this constraint and hence are better options, namely [ba.ta] and [ba]. These candidates satisfy NoCoda, but violate Parse one or more times because one or more segments are deleted. In [ba.ta], for example, the final segment [k] is deleted. Nevertheless, this candidate is the optimal output (indicated by the pointing hand) because it satisfies the total set of constraints best. So the grouping structure [ba.ta] is the optimal way to group speech segments into syllables. This structure satisfies the stronger constraint NoCoda, whereas it only violates the weaker constraint Parse once. No other candidate yields better results with respect to these two constraints. This example shows that a form can be the best form for a given input even if it violates one of the constraints on linguistic forms.

If the two constraints NoCoda and Parse were ordered the other way around, with Parse being stronger than NoCoda, the optimal form would be [ba.tak]. This is the effect of deletion now being worse than having a syllable with a coda. A general view in OT holds that languages are characterized by the same set of constraints and that differences among languages arise as the result of constraint reranking. If the linguistic constraints are identified, the OT model yields very precize predictions with respect to possible and impossible structures and meanings across languages.

Although Prince and Smolensky introduced their theory as a theory of phonology, their optimization approach spread to other linguistic disciplines such as syntax (Bresnan, 2000; Grimshaw, 1997), semantics and pragmatics (Blutner, 2000; Hendriks & de Hoop, 2001) as well. In general, OT seems to offer a fruitful way to investigate many cognitive phenomena that exhibit tendencies rather than clear-cut distinctions. The system of violable output-oriented constraints in OT thus allows us to formalize Lerdahl and Jackendoff’s system of preference rules. Importantly, OT incorporates a well-defined mechanism of conflict resolution, yielding clear predictions in the case of conflicting tendencies. Gilbers and Schreuder (2002) already noted the similarity between OT and Lerdahl and Jackendoff’s system, and show how OT is able to provide an analysis of the metrical structure of music. In our article, we will focus on the grouping structure of music. In particular, we will investigate the possibility of a coherent OT model of grouping in music. To this purpose, Lerdahl and Jackendoff’s theory will be translated into an OT framework. The OT constraints will then be implemented in a computational model, which allows us to compare the effects of different orderings of the same set of constraints.

3. Musical grouping

In order to describe the process of grouping in music, Lerdahl and Jackendoff (1983) distinguish two kinds of rules: grouping well-formedness rules (GWFR’s) and grouping preference rules (GPR’s). GWFR’s cannot be violated and define all possible grouping structures. GPR’s may be violated and define preferred grouping structures. We can distinguish two kinds of GPR’s: those acting on the first level (the actual pitch events), and those acting on groups. In this article, we will limit ourselves to rules of the first kind.

The properties of a series of notes can be such that application of the preference rules results in several mutually incompatible grouping structures. When preference rules are in conflict, nothing in Lerdahl and Jackendoff's system tells us which of the resulting grouping structures is the correct one. Lerdahl and Jackendoff argue that this is exactly what we want because listeners experience ambiguity in these cases. According to Clarke (1986:9), however, this indeterminacy of application within the preference rule system "severely undermines the explicitness and precision of the theory". Moreover, the different grouping structures that result from a conflict between rules do not seem to be equally plausible. Some of the preferences appear to be stronger than others.

The two broad Gestalt principles underlying the preference rules, namely proximity and similarity, were borrowed from the domain of vision. In the domain of vision, conflicts between these two principles have already been investigated (Quinlan & Wilton, 1998). With respect to musical perception, Tenney and Polansky (1980) emphasize the need to determine how to weigh the different parameters involved in proximity and similarity relative to each other. As we aim to demonstrate in this article, the linguistic theory OT offers a principled way to resolve conflicts between the preference rules concerning musical grouping by viewing musical perception as a process of optimization over a set of constraints of different strength.

Lerdahl and Jackendoff's preference rules can be reformulated as violable OT constraints as follows:

SINGLES (GPR 1): Groups never contain a single element.

PROXIMITY SLUR/REST (GPR 2a): No group contains a contiguous sequence of three notes, such that the interval of time from the end of the second note to the beginning of the third is greater than that from the end of the first note to the beginning of the second.

PROXIMITY ATTACKPOINTS (GPR 2b): No group contains a contiguous sequence of three notes, such that the interval of time between the attackpoints of the second and the third note is greater than that between the attackpoints of the first and the second note.

CHANGE REGISTER (GPR 3a): No group contains a contiguous sequence of three notes, such that the interval from the second note to the third is bigger than that from the first note to the second.

CHANGE DYNAMICS (GPR 3b): No group contains a contiguous sequence of three notes, such that the first two share the same dynamics, different from the third.

CHANGE ARTICULATION (GPR 3c): No group contains a contiguous sequence of three notes, such that the first two share the same articulation, different from the third.

CHANGE LENGTH (GPR 3d): No group contains a contiguous sequence of three notes, such that the first two share the same length, different from the third.

Constraints in OT must be formulated as general as possible. Also, OT constraints are categorical ('candidate A is X' or 'candidate A is not X'). Gradient constraints ('candidate A is better than candidate B') are not allowed. Therefore, the preference rules are formulated in terms of forbidden groups, rather than preferred boundaries between groups. As we will show in section 5.2, the interaction among these constraints allows the preferred boundaries to be determined.

In their formulation of the preference rules, Lerdahl and Jackendoff considered sequences of four notes. Their preference rules then decided whether or not a boundary should be placed between the second and the third note. To allow for incremental parsing (i.e., on a note-by-note basis as more of the input becomes available), however, we omit reference to the fourth note. It is widely assumed in the psycholinguistic literature that humans process language in an incremental fashion (i.e., on a word-by-word basis without delay). As soon as a word is encountered, it is integrated into a syntactic representation of the sentence (Steedman, 1989; Sturt & Crocker, 1996). Similarly, a musical percept also seems to be built up incrementally while listening to an unfolding tone series (Povel & Jansen, 2001). By working with a three-note window rather than a four-note window, our formulation deviates from the original formulation of the rules, but not from the Gestalt principles underlying these rules. Similarity is still an important principle in the reformulated constraints, but only between notes within the same group (i.e., the first two notes of a sequence of three notes), not between two groups (such as between the group consisting of the first two notes and the group consisting of the last two notes in a sequence of four notes). We assume the latter type of similarity to be regulated by higher-level preference rules or constraints, which we do not discuss in this article.

GPR 1 states that a grouping structure consisting of small groups should be avoided: very small-scale grouping perceptions tend to be marginal. GPR 2 defines the effect of temporal proximity in music. Proximate notes (e.g., slurred notes) should ideally be assigned to the same group. GPR 3 formalizes the intuition that notes with the same properties are grouped. Some of these rules pertain to the same properties of notes. For instance, in many cases where GPR 2b is violated, GPR 3d will be violated too because they both consider the length of the notes. Only when the notes are separated by a rest, GPR 2b might be violated (given that it makes a judgement based on the interval between attack-points) whereas GPR 3d remains unviolated (since it makes a judgement based on the actual length of the notes themselves). However, this overlap between the constraints is not problematic for OT.

The constraints evaluate the candidates (i.e., possible grouping structures), which are generated on the basis of a given input (a series of notes). It is important to realize that, in accordance with OT, all constraints are assumed to apply to all candidates. This differs from hypothesizing that listeners choose which rule to apply in which case, as Deliège (1987) does in her experimental study, which was also designed to test Lerdahl and Jackendoff's preference rules. The proposed constraints do differ in their strength, or relative importance. It is the strength of the constraints that we are concerned with in this article.

4. Computational OT model

On the basis of the constraints introduced in the previous section, we developed a computational model of musical grouping using the logic programming language Prolog (see van der Werf (2003) for details of the implementation). This is by no means the first parsing model that is based on Lerdahl and Jackendoff's GTTM. For example, Temperley (2001) already developed a rule-based computational model of rhythmic pattern recognition which was partly based on Lerdahl and Jackendoff's theory. However, our model is the first to use OT for the resolution of the conflicts that are inherent to Lerdahl and Jackendoff's grouping preference rules. An alternative, memory-based, approach to grouping that does not use any predefined rules or constraints is presented by Bod (2002).

Besides a Prolog translation of the constraints, our computational model (parser) includes an implementation of the OT processes GEN (which generates the candidate outputs) and EVAL (which evaluates these candidates by means of the ranked set of constraints). Due to the properties of Prolog, the set of candidate outputs is finite. Also, all candidates conform to Lerdahl and Jackendoff's grouping well-formedness rules. Notes are represented as an ordered list of ontime (in ms), offtime (in ms), frequency (in Hz), and dynamics (dynamical mark from ppp to fff). For example, the string n[200,450,440,mf] implements a 250 millisecond note of 440 Hertz with the dynamic mark mezzoforte. Because no information about articulation was included in the representation of notes, we did not implement the grouping rule referring to change of articulation (GPR 3c). However, if we would have added an articulation feature to our representation of notes, we could have implemented this constraint as well.

The only parameter to be set was the hierarchical ranking of the constraints. Given a particular ranking of the constraints, our computational model is able to assign a preferred grouping structure to any sequence of notes. To arrive at a plausible ranking of these constraints, we performed an experiment with a small number of subjects, which is discussed in the next section.

5. Testing the model

Lerdahl and Jackendoff (1983) do not report testing their rules with actual subjects (but see Deliège (1987) for an attempt to experimentally confirm Lerdahl and Jackendoff's rules). In order to investigate the explanatory power of our model and to determine the relative importance of the constraints, a listener study was performed with ten subjects. Our assumption was that Lerdahl and Jackendoff's rules and the corresponding OT constraints are correct generalizations with respect to the factors influencing the way human listeners group notes in music. On the basis of this assumption, we hypothesize that there is a particular hierarchical ranking of the constraints which explains the grouping structures selected by human listeners for particular sequences of notes. Thus the main goal of the experiment is to arrive at an empirical plausible hierarchy of the constraints GPR 1 - GPR 3d.

5.1 Methods

After a short introduction each subject was presented with 20 recordings of musical phrases, 5 notes in length, together with the same phrases in printed score. Every recording was played twice. The stimulus on paper contained no measures nor indication of time in order to avoid all possible grouping cues other than the notes themselves. The audio fragment was presented with a headphone and played at an appropriate level so that all notes could easily be heard. Subjects were asked to group the notes on the printed scores by circles.

5.2 Stimuli

The stimuli used in the experiment were series of five notes in MIDI, in combination with a written score. The MIDI audio format includes ontime, offtime, pitch, instrument and intensity. The MIDI instrument chosen for the experiment was an ocarino, which has a more or less constant intensity and spectrum during the note. Each score of a stimulus was printed on a separate paper using Sibelius® music notation software. The pitches of the stimuli were taken from the Thema Regis from the Musical Offering by Johann Sebastian Bach (1685 – 1750).

We constructed our stimuli (see Appendix A) in such a way as to gain maximal insight into the hierarchical ranking of the constraints. To establish their ranking, 17 of the 20 stimuli were constructed in such a way that they would create a conflict between at least two constraints. When two constraints are in conflict, there is no way to satisfy both constraints. The relative ordering of the constraints determines which constraint must be violated in order to satisfy the other (stronger) one. Consequently, the chosen grouping structure (i.e., the optimal candidate) provides insight into the relative ordering of the conflicting constraints. For each stimulus, 16 different grouping structures are possible. Appendix B lists the constraint violations for one of these structures, namely the most given response. Other possible grouping structures for a given stimulus violate other constraints. In order to test the empirical validity of the constraints themselves, we also constructed 3 stimuli (A, D, and U) for which a grouping structure was possible that satisfied all constraints.

As an illustration of the conflicts between the proposed constraints, consider the following stimulus:

Figure 2. Stimulus E.

The interaction among constraints can be depicted in the following OT tableau:

Stimulus E	GPR 1	GPR 2a	GPR 3b
2+3		*	
3+2			*
2+2+1	*		

Figure 3. An OT tableau for stimulus E.

The input (here: stimulus E) and a few of the candidate outputs (possible grouping structures for stimulus E) are listed in the first column. Grouping structures are represented by the number of notes in each group, separated by a plus sign. An asterisk indicates a constraint violation. The relevant constraints are listed in order of descending strength from left to right across the first row (cf. OT). Of course, as in OT, all constraints are assumed to apply in all cases. However, the omitted constraints (such as GRP 2b) are not violated by any of the candidates shown here.

In stimulus E, GPR’s 1, 2a and 3b are in conflict. They cannot be satisfied at the same time. If grouping structure 2+3 is chosen in order to satisfy GPR 3b and GPR 1, GPR 2a is violated. Vice versa, if grouping structure 3+2 is chosen in order to satisfy GPR 1 and GPR 2a, GPR 3b is violated. A solution might be to group the notes as 2+2+1, but then GPR 1 is violated. If subjects nevertheless choose grouping structure 2+2+1, then GPR 1 must be the weakest of the three constraints because violating this constraint is preferred to violating any of the other two constraints. By determining the relative ordering of all pairs of constraints based on a conflict between these constraints, we can establish the total hierarchy of constraints. Because in OT no number of violations of weaker constraints can compensate for the violation of a stronger constraint, the predictions of our model differ from the predictions of a numerical model such as the one discussed by Tenney and Polansky (1980). In particular, our model is more restricted and hence yields better circumscribed predictions.

Note that, although the constraints are formulated in terms of forbidden groups, the candidates are representations of the possible grouping structures for a given stimulus, including an indication of the position of the boundary or boundaries. As a result, the preferred position of the boundary follows from the interaction between the constraints. For example, if GRP 3b is the weakest constraint in our example, grouping structure 3+2 is the preferred structure. Consequently, the preferred boundary in stimulus E immediately follows note 3.

5.3 Subjects

10 subjects with intermediate musical experience (no professional musicians) were asked to participate in the experiment. The average period the subjects had been playing an instrument was 16.7 years, and the average period they had taken lessons was 10.5 years. No subject reported having problems with hearing.

6. Results

The results are given in figure 4. Stimulus J is left out corresponding to musical convention. The first column states the name of the stimulus. The second column gives the grouping structure that was chosen more often than any other grouping structure for that stimulus. Column three gives the number of subjects (denoted by k) that selected this most given response. The last column lists the probability that the number of subjects mentioned in the previous column based their judgements on chance. We will discuss these probabilities in the next section.

Stimulus	**Most given response**	**Subjects (k)**	**P (at least k)**
A	2+3	6	1.00×10^{-5}
B	2+2+1 / 2+3	4	0.00236
C	2+3	6	1.00×10^{-5}
D	2+3	7	3.78×10^{-7}
E	2+2+1	6	1.00×10^{-5}
F	2+3	8	9.35×10^{-9}
G	2+3	5	0.000184
H	2+3	7	3.78×10^{-7}
I	2+2+1	6	1.00×10^{-5}
K	3+2	5	0.000184

L	2+2+1	6	1.00×10^{-5}
M	2+3	6	1.00×10^{-5}
N	2+3	8	9.35×10^{-9}
O	4+1 / 3+2	3	0.0210
P	4+1	5	0.000184
Q	2+2+1	3	0.0210
R	3+2	4	0.00236
S	2+3	3	0.0210
T	3+2	6	1.00×10^{-5}
U	3+2	3	0.0210

Figure 4. Most given response per stimulus (N=10).

In a number of cases, two different responses were chosen the same number of times, and more often than other responses. In that case, we included both groupings in the table, separated by a slash.

7. Discussion

The last column in figure 4 lists the probabilities for at least k subjects out of N subjects to make the same decision out of 16 different grouping structures on the basis of chance (see Van der Werf & Hendriks (2004) for details). As can be seen here, these probabilities are extremely small. From this we may conclude that the subjects' responses on the stimuli are based on certain preferences, and are unlikely to be explained through pure chance. By considering our computational model in more detail, we might be able to provide an answer to the question why subjects display these preferences.

7.1 Validity of the constraints

Stimuli A, D, and U were included in our experiment to test the validity of the constraints themselves. If the proposed constraints are correct and if no other constraints play a role in musical grouping, we expect all subjects to give the response 2+3, which is the optimal structure for these three stimuli because it satisfies all constraints. However, from figure 4 it can be seen that this is not the case. Although subjects indeed showed a strong preference for the optimal candidate in stimuli A and D, subjects did not agree upon the preferred structure for stimulus U. Also, the preferred structure for this stimulus incurs a constraint violation, namely of GPR 3d. This suggests that either the constraints as they are formulated here do not accurately express the correct generalizations, or some additional as yet unknown factor might be involved here. As a matter of fact, Deliège (1987) in her experimental study finds evidence for the necessity of introducing extra rules. However, more research with a larger pool of subjects is needed to decide on this issue. With respect to testing the constraints themselves, a complicating factor is that in OT all constraints apply to all candidates. As a result, their validity cannot be determined by studying the constraints in isolation but must be determined on the basis of their interaction with other constraints.

7.2 Group results

To determine the constraint ranking that explains the group results best, we looked at the most given responses for each stimulus. The constraint ranking can be determined by looking at stimuli that are constructed based on conflicting constraints. If only responses given by more than half of the subjects (i.e, k > 5) are considered, a consistent but incomplete constraint hierarchy is obtained, where A >> B means that constraint A is stronger than constraint B:

{PROXIMITY SLUR/REST (GPR 2a),
CHANGE DYNAMICS (GPR 3b),
CHANGE LENGTH (GPR 3d)}
>> SINGLES (GPR 1)
>> PROXIMITY ATTACKPOINTS (GPR 2b)
>> CHANGE REGISTER (GPR 3a)

The constraints GPR 2a, GPR 3b and GPR 3d are strongest, but unranked with respect to each other. The relative ranking of these three constraints is obtained by also taking into account stimulus S (k=3). This stimulus provides weak evidence that GPR 3d is stronger than GPR 3b. Furthermore, GPR 2a is never violated (with the exception of the second most given response for stimulus B), while GPR 3d and GPR 3b are. Based on these observations, we assume the following ranking among these three constraints:

PROXIMITY SLUR/REST (GPR 2a)
>> CHANGE LENGTH (GPR 3d)
>> CHANGE DYNAMICS (GPR 3b)

Rather unexpected are the low positions of GPR 1 (SINGLES) and GPR 3a (CHANGE REGISTER). The low position of GPR 1 might be partly due to a bias against reproducing the same response (in our experiment, 2+3 or 3+2) over and over again. Note that the candidates 2+3 and 3+2 are the only two candidates that obey GPR 1. The low position of GPR 3a is also surprizing. At first sight, change of pitch is an important indication that a new group should be started. However, this does not show from the data. This constraint was violated in the most given response to several stimuli (stimuli G, K, M, N and T, see Appendix B). As a result, the group model that best fit the data involves a ranking in which GPR 3a is placed lowest. A reason might be that we used sequences of only five notes, whereas melody often consists of more notes. Because GPR 3a is the only constraint concerned with melodic structure, more elaborate passages might yield a different ranking of this constraint.

Deliège (1987), in two listener studies, also provides evidence for the validity of Lerdahl and Jackendoff's grouping preference rules. Moreover, her work suggests a relative salience scale for the rules. Unfortunately, her results are not directly comparable with ours since her experimental setup and basic assumptions differ substantially. For example, Deliège (fig. 4, p.329) claims that groups with one sound are never preferred. Hence, she does not include this possibility in her study. However, we did include GPR 1 in our constraint set and found that groups consisting of one sound can be preferred in certain cases. That is, GPR 1 seems to be a violable constraint. The omission of GPR 1 may have affected Deliège's interpretation of the results. Also, Deliège uses the original four-note window formulation of the preference rules, whereas we used a three-note window in our formulation of the preference rules. As a result, Deliège's rules apply in a more restricted set of circumstances than our constraints. A third difference is that in the second listener study, in which Deliège aimed to investigate cases of conflict between two preference rules, subjects were allowed to impose one segmentation only for each sequence of nine sounds. As a result, the created conflict is not between two rules (because in many cases they could be applied both, resulting in two segmentations, perhaps violating GPR 1), but between the rules on the one hand and the restriction on segmentation on the other. Due to this restriction, subjects could not apply all rules simultaneously, but had to choose which rule to apply. This then gives us an indication of the relative salience of the rules, or of the relative salience of the acoustic and temporal properties the rules pertain to. However, relative salience may not be identical to relative strength, which we investigated in this study. Rather, it may be a side-effect of the dynamics of the process of musical perception, as is also suggested by the recency effects Deliège finds.

7.3 Intersubject variation

Our results show a moderate amount of variation among the subjects. To provide a measure of the difference between two results, we introduce the notion of closeness. The closeness C of two equally long vectors *a* and *b*, both of length m, is a measure of the similarity of the two vectors, and ranges from 0 (completely different) to 1 (completely identical) (see Van der Werf (2003) and Van der Werf & Hendriks (2004) for computational details). Applied to our results, closeness is taken to be a measure of the similarity between two results on the same set of stimuli. The average closeness between individual subjects and the group results (obtained by taking the most chosen responses in the experiment) is 0.57 ± 0.20. In other words, there is some variation among the subjects.

We modelled the group results as well as the results of individual subjects by determining the constraint ranking that best fit the responses. Because we used a computational model, this could be done straightforwardly. The ranking corresponding to the group results was already given in section 7.2. The closeness between the computational group model and the experimental group results is 0.7 (with closeness ranging from 0.10 onward because we defined group results in terms of the most given response). This means that the computational model does not predict the experimental group results with complete accuracy. Nevertheless, the closeness between the computational group model and the experimental group results (0.7) is at least as large as the closeness between the individual subjects and the experimental group results (0.57 ± 0.20). That is, the computational group model behaves as an above average subject.

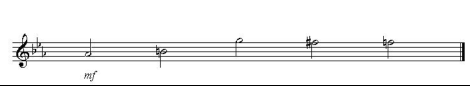
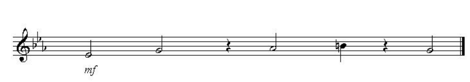
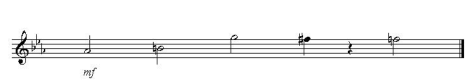
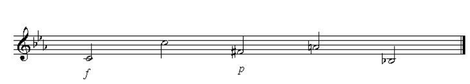
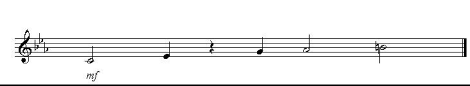
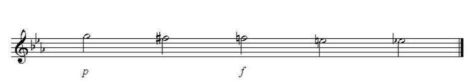
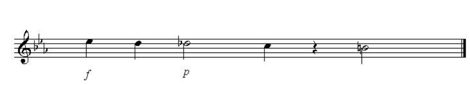
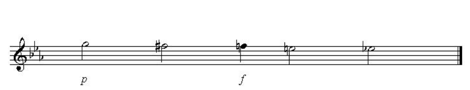
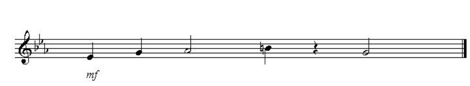
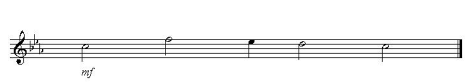
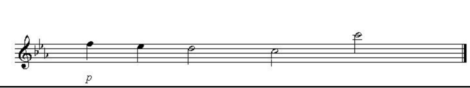
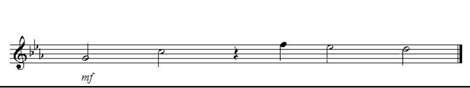
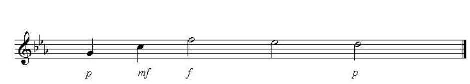
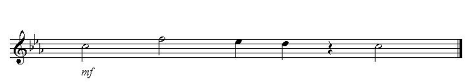
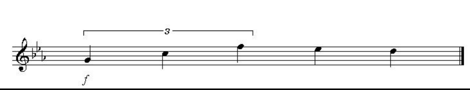
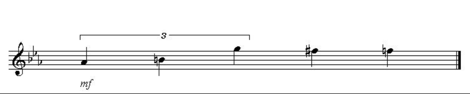
By reranking the constraints, separate computational models could be made for individual subjects. The average closeness between each computational subject model and the corresponding human subject is 0.59 ± 0.13. Because this does not substantially differ from the closeness between the computational group model and the experimental group results (0.7), this suggests that individual differences cannot be explained by constraint reranking. In OT, it is assumed that differences among languages can be explained by reranking the same set of constraints. Apparently, individual differences in musical grouping among people from the same cultural group should not be compared to differences between speakers of different languages.

8. Conclusions

We used the mechanisms of Optimality Theory as the basis of our musical parser. This resulted in a comprehensive and working computational model of musical grouping. When we tested the model experimentally, we did not find any evidence suggesting that we should abandon the main assumptions that we started out with: 1) parsing a musical surface is not a coincidental process, but is governed by constraints, 2) these constraints take the form of preferences that can be violated, and 3) these constraints differ in strength, or relative importance. Using our computational model, we derived a ranking (according to strength) of the constraints on musical grouping that best fit the data.

Whereas language is primarily concerned with communication, music is primarily concerned with expression. However, language and music both achieve their goals in the same way, namely through highly structured sound patterns. In order to understand these patterns, listeners of music and language are faced with the same problem of uncovering the underlying structure. In this article we showed that this process, which is fundamental to both domains of cognition, could be modelled using the same optimization mechanism.

Appendix A: Stimuli

A		L	
B		M	
C		N	
D		O	
E		P	
F		Q	
G		R	
H		S	
I		T	
K		U	

Appendix B: Conflicting constraints for the most given response per stimulus

Stimulus	**Most given response**	**Violated constraints**	**Non-vacuously satisfied constraints**
A	2+3	-	GPR 1, GPR 3a
B	2+2+1	GPR 1	GPR 2a, GPR 3a, GPR 3d
	2+3	GPR 2a	GPR 1, GPR 3a, GPR 3d
C	2+3	GPR 2b	GPR 1, GPR 3a, GPR 3d
D	2+3	-	GPR 1, GPR 3b
E	2+2+1	GPR 1	GPR 2a, GPR 3b, GPR 3d
F	2+3	GPR 2b	GPR 1, GPR 3b, GPR 3d
G	2+3	GPR 2b, GPR 3a	GPR 1, GPR 3d
H	2+3	GPR 2b	GPR 1, GPR 2a, GPR 3d
I	2+2+1	GPR 1	GPR 2a, GPR 2b, GPR 3a, GPR 3d
K	3+2	GPR 3a	GPR 1, GPR 2b, GPR 3d
L	2+2+1	GPR 1	GPR 2a, GPR 2b, GPR 3a, GPR 3d
M	2+3	GPR 3a	GPR 1, GPR 3b

N	2+3	GPR 2b, GPR 3a	GPR 1, GPR 2a, GPR 3d
O	4+1	GPR 1, GPR 2b, GPR 3d	GPR 2a
	3+2	GPR 3d	GPR 1, GPR 2a, GPR 2b
P	4+1	GPR 1, GPR 2b, GPR 3b, GPR 3d	GPR 2a
Q	2+2+1	GPR 1	GPR 2a, GPR 2b, GPR 3a, GPR 3d
R	3+2	GPR 3d	GPR 1, GPR 2b, GPR 3a
S	2+3	GPR 3b	GPR 1, GPR 2b, GPR 3a, GPR 3d
T	3+2	GPR 3a	GPR 1, GPR 2b, GPR 3d
U	3+2	GPR 3d	GPR 1, GPR 2b

All constraints apply to all 16 possible grouping structures. For each stimulus, one of these grouping structures (namely the most given response) is given in column 2. Constraints that are violated by the most given response are listed in column 3. Constraints that are satisfied by the most given response although they could have been violated are listed in column 4. Constraints that are not mentioned for a given stimulus are satisfied vacuously (e.g., they cannot be violated in this particular case because the stimulus is such that the conditions under which a violation can occur do not hold).

Acknowledgments

The authors thank the two anonymous reviewers for their valuable comments. Furthermore, Ronald Zwaagstra is thanked for his help with the statistical analysis of the data. Petra Hendriks gratefully acknowledges the Netherlands Organisation for Scientific Research, NWO (grants no. 051.02.070 and 015.001.103).

References

Blutner, R. (2000). Some aspects of optimality in natural language interpretation. *Journal of Semantics*, 17, 189-216.

Bod, R. (2002). Memory-Based Models of Melodic Analysis: Challenging the Gestalt Principles. *Journal of New Music Research* 31.1, 27-37.

Bresnan, J. (2000). Optimal syntax. In: J. Dekkers, F. van der Leeuw & J. van de Weijer (eds.), *Optimality Theory: Phonology, syntax, and acquisition*. Oxford: Oxford University Press, 334-385.

Clarke, E.F. (1986). Theory, Analysis and the Psychology of Music: A Critical Evaluation of Lerdahl, F. and Jackendoff, R., A Generative Theory of Tonal Music. *Psychology of Music* 14, 3-16.

Deliège, I. (1987). Grouping Conditions in Listening to Music: An Approach to Lerdahl and Jackendoff's Grouping Preference Rules. *Music Perception* 4.4, 325-360.

Gilbers, D. & M. Schreuder (2002). Language and music in Optimality Theory. Unpublished manuscript, University of Groningen.

Grimshaw, J. (1997). Projections, heads, and optimality. *Linguistic Inquiry*, 28, 373-422.

Hendriks, P. & de Hoop, H. (2001). Optimality theoretic semantics. *Linguistics and Philosophy*, 24, 1-32.

Lerdahl, F. (2001). *Tonal Pitch Space*. Oxford: Oxford University Press.

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

Povel, D.-J. & Jansen, E. (2001). Perceptual Mechanisms in Music Processing. *Music Perception* 19.2, 169-198.

Prince, A. & Smolensky, P. (1993/2004). *Optimality Theory: Constraint interaction in generative grammar*. Oxford: Blackwell. (distributed as unpublished ms., Rutgers University, New Brunswick, and University of Colorado, Boulder, NJ, 1993)

Quinlan, P.T. & Wilton, R.N. (1998). Grouping by Proximity or Similarity? Competition Between the Gestalt Principles in Vision. *Perception* 27, 417-430.

Steedman, M. (1989). Grammar, interpretation and processing from the lexicon. In: W. Marslen-Wilson (ed.), *Lexical Representation and Process*. Cambridge, MA: MIT Press, 463-504.

Sturt, P. & M. Crocker (1996). Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes* 11.5, 449-494.

Temperley, D. (2001). *The Cognition of Basic Musical Structures*. Cambridge, Mass.: MIT Press.

Tenney, J. & L. Polansky (1980). Temporal Gestalt Perception in Music. *Journal of Music Theory* 24, 205-241.

Van der Werf, S. (2003). Grouping in Language and Music, two faces of the same problem? MA Thesis, Department of Artificial Intelligence, University of Groningen.

Van der Werf, S. & Hendriks, P. (2004). A constraint-based approach to grouping in language and music. In: R. Parncutt, A. Kessler & F. Zimmer (eds), *Proceedings of the Conference on Interdisciplinary Musicology (CIM04)*, Graz.