A constraint-based approach to grouping in language and music

Sybrand van der Werf
Academy of Theatre and Conservatorium Maastricht, The Netherlands
sybrand.v.d.werf@freeler.nl

Petra Hendriks
Centre for Language and Cognition Groningen, University of Groningen, The Netherlands
P.Hendriks@let.rug.nl www.let.rug.nl/~hendriks

In: R. Parncutt, A. Kessler & F. Zimmer (Eds.)
Proceedings of the Conference on Interdisciplinary Musicology (CIM04)
Graz/Austria, 15-18 April, 2004 http://gewi.uni-graz.at/~cim04/

Background in music psychology and music computation. In Lerdahl and Jackendoff's Generative Theory of Tonal Music (1983), the authors try to formalize cognitive processes seen in music, referring to Chomskyan (generative) linguistics. Among other processes found in musical cognition (such as harmonic tension and detecting rhythmic patterns), Lerdahl and Jackendoff describe the process of grouping single pitch-events into larger scale units. They propose a rule-based system involving two kinds of rules: well-formedness rules and preference rules. A rule-based computational model of rhythmic pattern recognition which was partly based on Lerdahl and Jackendoff's theory was developed by Temperley (2001).

Background in linguistics. A recently developed, but already quite influential theory in several subdomains of linguistics is Optimality Theory (Prince & Smolensky, 1993, 1997). A distinguishing property of Optimality Theory, which follows from its roots in neural network modeling, is its use of violable constraints instead of inviolable rules. The interaction among these violable constraints can be characterized formally, which allows for the computational modeling of linguistic processes in a straightforward manner. The resulting system is able to deal with preferences but at the same time yields very precize predictions with respect to possible and impossible linguistic structures and meanings.

Aims. We aim to develop a cognitively motivated musical parser that is based on constraint optimization in linguistics for grouping pitch-events into larger-scale units.

Methods. The rules proposed by Lerdahl and Jackendoff governing the process of grouping in music were implemented in the logic programming language Prolog. Their interaction is modelled as OT constraint interaction. In an experiment involving ten subjects the preference rules were evaluated. Using different techniques, the results for both individual subjects and for the group of subjects were compared with each other and with the results of the computational model (the musical parser).

Results. The rules proposed by Lerdahl and Jackendoff were implemented without any difficulties and resulted in a comprehensive and working OT based model. In contrast to Lerdahl and Jackendoff's theory, our computational model is also able to account for the interaction among the preference rules. Our model determines an optimal grouping structure for every given sequence of pitch-events. Although there was some variation among the subjects involved in the experiment, the model's output was comparable to the results of the subjects.

Conclusions. From a comparison of the results of our musical parser with the results of a number of experimental subjects, we conclude that language and music seem to make use of similar mechanisms for the grouping of auditory events. These mechanisms include the generation of possible grouping structures and the evaluation of these structures against a set of violable constraints. The optimal musical group is the one that optimally satisfies the total set of constraints on musical grouping.

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 grouping in music.

Linguistics and musicology

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 phrases. This division into smaller and smaller units continues until we are left with morphemes, which are the smallest units of meaning in a language. Also with respect to the metrical structure of language, hierarchical representations are assumed. Sound segments are organized into syllables, which unite into feet, which again unite into words and phrases. This hierarchical organization of language at different levels of linguistic structure 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, 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.

Optimality Theory and music

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, 1997). 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. As a result, an optimal output need not satisfy all constraints perfectly, but only needs to do so optimally. An output is optimal if no alternative does better, given the relative strength of the constraints. An optimal output might violate certain constraints, but only if there is no other way to avoid violation of a stronger constraint. Therefore, OT constraints are in principle violable. As a consequence, OT constraints express linguistic tendencies rather than unviolable principles.

Constraint interaction in OT

As an illustration of the interaction among constraints in OT, 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 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 pronounced. Pronunciation 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 symplified 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 descending order of 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. It 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 is the view 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 seems to 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 linguistic phenomena that exhibit tendencies rather than clear-cut distinctions. Thus the system of violable output-oriented constraints in OT seems well-suited to formalize Lerdahl and Jackendoff's system of preference rules. Gilbers and Schreuder (2002) already noted the similarity between OT and Lerdahl and Jackendoff's system, and show how OT might be able to provide an analysis of the metrical structure of music. In our paper, 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 OT model, which allows us to compare the effects of different orderings of the same set of constraints.

Musical grouping

In order to describe the process of grouping in music, Lerdahl and Jackendoff 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 perceptual possible grouping-structures. GPR's are soft and may be violated in order to satisfy other GPR's. They define a preferred structure, comparable with the optimal candidate in OT.

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 paper, 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 collide, 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. However, the different grouping structures that result from a collision of rules do not seem to be equally plausible. Moreover, some of the preferences appear to be stronger than others. OT offers a way to explain these observations by viewing conflict resolution as a process of optimization over a set of ranked constraints.

Soft constraints on musical grouping

To be able to translate Lerdahl and Jackendoff's preference rules into OT constraints, we modified them in two ways. First, the rules proposed by Lerdahl and Jackendoff define properties of preferred groups. To fit these rules into an OT framework, these rules had to be reformulated as violable constraints. In particular, it should be possible to determine whether the rule is violated or not. Secondly, for the application of preference rules Lerdahl and Jackendoff consider sequences of four notes. The preference rules then decide whether or not a boundary should be placed between the second and the third note. To allow for incremental parsing, we omitted reference to the fourth note. At the time a listener hears the third note, we assume that he or she makes a decision with respect to the hierarchical position of this note depending of the properties of the first three notes: does this note introduce a new group, or does this note belong to the same group as the second note? The properties of the fourth note should not influence this decision. For this reason, we omitted the fourth note from our OT formulation of Lerdahl and Jackendoff's rules. However, reference to a fourth note can easily be included in our formulation of the constraints.

The preference rules stated by Lerdahl and Jackendoff are given below in their modified OT format.

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.

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, or, stated in terms of boundaries, a boundary is placed between notes that differ with respect to their properties.

Computational OT model

Our aim was to develop a computational model of the interaction among the OT constraints on musical grouping. This model was developed in the logic programming language Prolog. Using lists in Prolog as a representation of groups, a number of properties of groups automatically follow. These properties (contiguity of elements, possibility of embeddedness and the impossibility of cross-reference are of interest to us) are exactly the properties expressed by the grouping well-formedness rules of Lerdahl and Jackendoff. This has a desired side-effect: the set of candidates becomes finite in size. In standard OT the set of candidates is assumed to be infinite in size, in order to prevent selection of candidates before the candidates are evaluated by means of the constraints.

The representation of notes used in our implementation was 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 440 Hz note of 250 milliseconds with the dynamic mark *mezzoforte*.

The OT constraints GPR 1 - GPR 3d as listed in the previous section were translated into Prolog. Prolog also allowed for a straightforward implementation of the OT routines GEN (which generates the candidate outputs) and EVAL (which evaluates these candidates by means of the ranked set of constraints). The only parameter to be set was the hierarchical ranking of the constraints. To arrive at a plausible ranking of these constraints, we performed an experiment with a small number of subjects, which will be reported on in the the next section. Given a particular ranking of the constraints, our computational model is able to assign a preferred grouping structure to any sequence of notes.

Because our implementation uses a representation of notes as 4-tuples of on-time, off-time, frequency and dynamics mark, it was impossible to implement the grouping rule referring to change of articulation (GPR 3c). Articulation (e.g., staccato, martelé or portato) is a very complex feature, which is often virtually indistinguishable in the audio signal. Because articulation in many cases is not marked in a score but left to the performer, we left the corresponding constraint out of consideration. However, if we would have added an articulation feature to our representation of notes, we could have implemented this constraint as well.

Testing the model

Lerdahl and Jackendoff do not report the testing of their rules with actual subjects. In order to test the psychological reality of these rules and to determine their relative importance, an experiment was performed with a set of 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 would explain the grouping structures selected by human listeners for particular sequences of notes. Thus the main aim of the experiment is to arrive at an empirical plausible hierarchy of the constraints GPR 1 - GPR 3d.

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.

Stimuli

The stimuli used in the experiment were series of five notes in MIDI in combination with a written score. MIDI is an audio format that includes ontime, offtime, pitch, instrument and intensity. The MIDI instrument that was chosen for the experiment is an ocarino (a small flute made

of pottery, originating from Italy). This instrument 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 (see Appendix). 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 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 two constraints. When two constraints are in conflict, the relative ordering of the constraints can be determined on the basis of the optimal candidate. If the optimal candidate violates constraint A and satisfies constraint B, if a suboptimal candidate satisfies constraint A and violates constraint B, and if these two candidates behave the same with respect to all other constraints, then constraint B must be stronger than constraint A. Evidently, in this case it is more important to satisfy constraint B than to satisfy constraint A. Finally, in order to test the empirical correctness of the constraints themselves, we also constructed 3 stimuli (A, D, and U) for which a grouping structure was possible that satisfied all constraints.

An example of an actual stimulus will show how this process works:

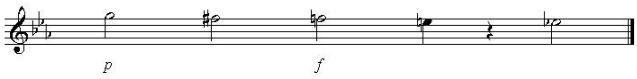


Figure 2. Stimulus E.

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

Figure 3. An OT tableau for stimulus E.

The notation for grouping structure used in figure 3 is the number of notes in each group, separated by a plus sign. In stimulus E GPR's 1, 2a and 3b are in conflict. If the grouping structure 2+3 is chosen in order to satisfy the constraint CHANGE DYNAMICS (GPR 3b), constraint PROXIMITY SLUR/REST (GPR 2a) is violated. Vice versa, if the grouping structure 3+2 is chosen in order to satisfy the constraint PROXIMITY SLUR/REST, constraint CHANGE DYNAMICS is violated. A solution might be to group the notes as 2+2+1, but then the constraint SINGLES (GPR 1) is violated. If we can determine the relative ordering of all pairs of constraints, based on a conflict between the two constraints of a pair, we can establish the total hierarchy of constraints.

Subjects

10 subjects with intermediate musical experience (no professional musicians) were asked to participate in the experiment. This is a small pool and further research should be done with a larger pool. 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.

Results

The results are given in figure 4 (on the next page). 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 that gave this response, denoted by k. The last column gives 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.

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. For example, stimulus O is grouped by 3 subjects as 4+1 and by 3 other subjecs as 3+2. Never given responses were 1+2+2, 3+1+1, 1+3+1, 1+2+1+1 and 1+1+1+2. The most given response overall was 2+3.

Stimuli	Most given response	Subjects (k)	P(at least k)
A	2+3	6	$1{,}00 \times 10^{-5}$
B	2+2+1 / 2+3	4	0,00236
C	2+3	6	$1{,}00 \times 10^{-5}$
D	2+3	7	$3{,}78 \times 10^{-7}$
E	2+2+1	6	$1{,}00 \times 10^{-5}$
F	2+3	8	$9{,}35 \times 10^{-9}$
G	2+3	5	0,000184
H	2+3	7	$3{,}78 \times 10^{-7}$
I	2+2+1	6	$1{,}00 \times 10^{-5}$
K	3+2	5	0,000184
L	2+2+1	6	$1{,}00 \times 10^{-5}$
M	2+3	6	$1{,}00 \times 10^{-5}$
N	2+3	8	$9{,}35 \times 10^{-9}$
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	$1{,}00 \times 10^{-5}$
U	3+2	3	0,0210

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

Discussion

The probability 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 is given by the following equation:

$$P = \sum_{k}^{N} \binom{k}{N} \cdot \left(\frac{1}{16}\right)^{k} \cdot \left(\frac{15}{16}\right)^{(N-k)}$$

Equation 1. Probability for at least k subjects out of N to make the same decision.

The last column in figure 5 gives the probabilities per stimulus for the given number of subjects (k). As can be seen, these probabilities are extremely small. From this we may conclude that the subjects' responses on the stimuli in the experiment are based on certain preferences and are unlikely to be explained through pure chance.

Constraints. Stimuli A, D, and U were included in our experiment to test the empirical correctness 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 in all 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. This indicates 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. However, more research with a larger pool of subjects is needed to decide on this issue.

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 are considered, a consistent but incomplete constraint hierarchy is obtained:

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

Figure 5. Incomplete constraint hierarchy resulting from considering only responses for which holds that k > 5.

The relative ranking of the three strongest constraints is obtained by also taking into account stimulus S (k=3). This stimulus provides weak evidence that CHANGE LENGTH is stronger than CHANGE DYNAMICS. Furthermore, PROXIMITY SLUR/REST is violated nowhere, while CHANGE LENGTH and CHANGE DYNAMICS are. Based on these observations, we assume PROXIMITY SLUR/REST to occupy the top position in our hierarchy:

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

Figure 6. Relative ranking of the three strongest constraints when stimulus S is also taken into account.

The resulting hierarchy has some unexpected features. First, the low position of SINGLES seems counter-intuitive. The term *group* almost always refers to more than one element. However, in the experiment SINGLES is frequently violated and ends as low as the 4th position. This might be 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 SINGLES.

Another conspicuous aspect of the hierarchy is the low position of CHANGE REGISTER. At first sight, change of pitch is an important indication that a new group should be started. Apparently this does not show from the data. A reason for this might be that the stimuli are constructed, instead of being existing musical fragments. Melodic pattern is not as full-fledged as it is in real music, where melody often consists of more than five notes. More elaborate passages might give a different ranking of this constraint because it is the only constraint concerned with melodic structure.

Intersubject results. Our results show a certain amount of variation among the subjects. To provide a measure of the difference between two results on the same set of stimuli, we introduce the notion of closeness. The closeness of two vectors **a** and **b**, both of length m, is defined as follows:

$$C(\mathbf{a},\mathbf{b}) \equiv \left(\frac{\sum_{i}^{m} \delta(a_i, b_i)}{m} \right)$$

Equation 2. Definition of closeness C of two vectors.

The altered delta function $\delta(a,b)$ gives 1 when a=b and 0 when $a \neq b$. Closeness ranges from 0 (completely different) to 1 (completely identical) and thus is a measure of the similarity between the results of individual subjects on the stimuli in our experiment and the results of the entire group on the same stimuli (obtained by taking the most chosen responses). The average closeness between individual subjects and group results is 0.57 ± 0.20. In other words, there is some variation among the subjects. A few of the subjects showed deviating responses (indicated by a low closeness), but most of the subjects behaved more or less similarly.

Closeness between the group model and the group results in our experiment is at least 0.10 because of the way we defined the group results in terms of the most given response. The closeness between the group model and the group results is 0.7. This means that the group model did not predict the group results with complete accuracy. Nevertheless, the responses generated by the group model were at least as close to the most given response (namely 0.7) as the responses by the individual subjects (0.57 ± 0.20). In other words, the group model behaved as an above average subject. Its responses were more similar to the group response than the responses of most human subjects in the experiment.

By reranking the constraints, separate models can be made for individual responses. Because the constraints were implemented in a computer model, this can be done straightforwardly. The closeness between these individual models and the corresponding individual responses can again be given in terms of their closeness. The average closeness between each model and the corresponding subject is 0.59 ± 0.13. Because the closeness between the individual models and the corresponding results (0.59 ± 0.13) does not substantially differ from the closeness between the group model and the 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.

Remaining issues

Our computer model was based on an almost direct translation of the preference rules of Lerdahl and Jackendoff into OT. However, it is possible that a different choice of preference rules or a different formulation of these rules yields better results.

An approach that might be worthwile pursuing is to encode the constraints in a local rather than global way (cf. Hammond, 1997). In our research, we focused on groups rather than on the boundaries between groups. The constraints in our computational model are constraints which promote or prohibit groups with a given structure. Also, in our experiment we asked subjects to circle notes that they felt to belong to the same group. However, it is computationally more attractive to have constraints referring to local properties of notes rather than to global properties of groups. Thus an interesting alternative formulation of the rules would be in terms of boundaries between groups rather than in terms of properties of groups. Of each note, we would only have to determine the structural position within a group: left edge, right edge or no edge. This type of local encoding drastically reduces the number of candidates that have to be considered. This is especially advantageous if larger musical fragments must be parsed.

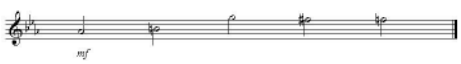
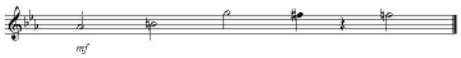
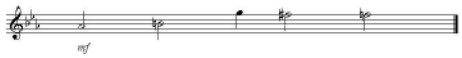
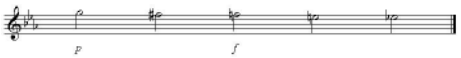
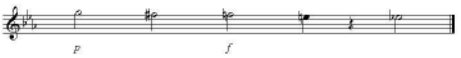
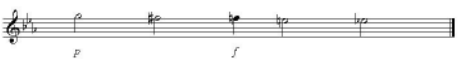
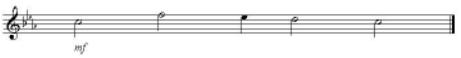
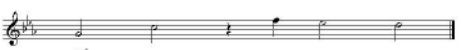
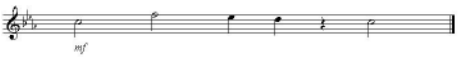
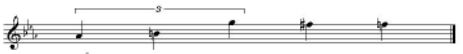
Our results suggest that grouping in music might be explained best by means of a system of violable constraints, as has also been argued for in linguistics. An interesting question is whether the constraints in our OT model for musical grouping correspond to similar constraints in language. Many of the features to which Lerdahl and Jackendoff's first-level preference rules and the corresponding OT constraints refer indeed have correlates in language which are used as diagnostics for phrasal boundaries. The features used by Lerdahl and Jackendoff include pauses between notes (GPR 2a), tempo (GPR 2b), pitch differences (GPR 3a), differences in loudness (GPR 3b), differences in articulation (GPR 3c), and differences in length (GPR 3d). In language, phrasal boundaries can be marked by pauses between phrases. Furthermore, speech sounds before a boundary are pronounced slower than speech sounds following a boundary. Also, the pitch and loudness of an utterance usually change after a boundary. Finally, pitch, loudness, syllable lengthening and richer articulation (which are all correlates of stress) can be used to mark the beginning or ending of a linguistic phrase. However, these prosodic cues do not fully determine the hierarchical structure of a linguistic utterance. In language, in addition to prosodic cues, there are syntactic constraints on how words can be combined into phrases and how phrases can be combined into larger phrases and sentences. In contrast, in music in principle any note may be combined with any other note. So although the features involved in musical grouping seem to be among the features used in linguistic grouping, their role is somewhat different. In music, as opposed to language, these features are the sole source of information concerning the grouping structure. As a consequence, we expect linguistic constraints on grouping to be similar but not identical.

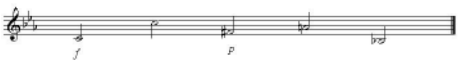
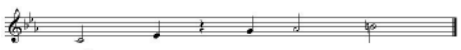
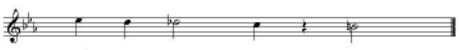
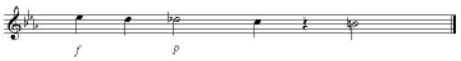
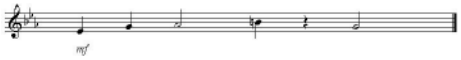
Conclusions

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 can be violated, and 3) these constraints differ in strength.

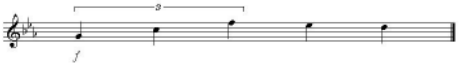
Music and language are different facets of human behaviour. Language is primarily concerned with communication, music with expression. But the means by which the two achieve their goals is essentially the same: highly structured sound patterns. In order to understand this sound, listeners of both music and language are faced with the same problem: uncovering the underlying structure. It was shown that this process, which is fundamental to both domains of cognition, could be modelled by using the same techniques.

Appendix: stimuli

A

B

C

D

E

F

G

H

I

K

L

M

N

O

P

Q

R

S

T

U


Acknowledgments

The authors would like to thank Ronald Zwaagstra 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.

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.

Gilbers, D., & M. Schreuder (2002). *Language and music in Optimality Theory.* Unpublished manuscript, University of Groningen, also available as ROA # 517-0103.

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

Hammond, M. (1997). *Parsing syllables: Modeling OT computationally.* Unpublished manuscript, University of Arizona, also available as ROA # 222-1097.

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

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

Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar.* Manuscript, Rutgers University, New Brunswick, and University of Colorado, Boulder, NJ. Also available as ROA # 537-0802.

Prince, A., & Smolensky, P. (1997). Optimality Theory: From Neural Networks to Universal Grammar. *Science* 275, 1604-1610.

Rutgers Optimality Archive (ROA), Rutgers University, New Brunswick. ROA is a distribution point for research in Optimality Theory. http://roa.rutgers.edu/

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

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.