Ranking Metrical Constraints in Music Perception – an Empirical Study

Mihailo Antovic

Faculty of Philosophy

University of Nis, Serbia

Correspondence:

Mihailo Antovic,

Department of English,

Faculty of Philosophy, University of Nis,

Cirila i Metodija 2, 18000 Nis, Serbia

mantovic@gmail.com

Abstract

This paper presents the results of an empirical research into the ordering of metrical preference rules/constraints from the group MPR5, as proposed in *A Generative Theory of Tonal Music* (Lerdahl & Jackendoff, 1983). A hundred and twenty randomly selected undergraduate students (30 musicians and 90 nonmusicians) were played twelve metrical sequences accorded with the MPR5 examples from GTTM , of which half complied with and half contradicted their expectancies. The participants were prompted to press a button when certain they heard a stressed beat. The answer distributions suggest that six constraints can be ranked into three larger groups, as follows: *physical stress* (dynamic, harmony), *melodic stress* (pitch, slur, length), *ornamental stress* (articulation). Musicians achieved better results than nonmusicians, and the response latencies considerably rose in the stimuli contradicting expectancies, but the internal constraint rankings seemed relatively stable irrespective of the two factors (musician/nonmusician; expected/unexpected suggestion).

Ranking Metrical Constraints in Music Perception – an Empirical Study

Introduction

In psycholinguistics and much of cognitive science the final quarter of the twentieth century was marked by gradual renunciation of strict binary choices in favor of relative preferences among a number of possible options. Originating from the well-known Gestalt principles of perception (e.g. Wertheimer, 1923), these preferential choices came to be called differently in different disciplines of cognitive science: in early pragmatics, they were named *conversational implicatures* (e.g. Grice, 1975), in music perception they became known as *preference rules* (Lerdahl & Jackendoff, 1983, hereinafter: GTTM), while in more recent linguistic and broader cognitive contexts they are often referred to as *constraints* (originally Ross, 1970; today Gilbers & Schreuder, 2002; Jackendoff, 2002). Regardless of the name, the construct implies that temporal structures are always parsed based on a set of physical changes in the quality of the stimulus, which are then perceived as clues as to where to divide the structure into meaningful wholes. The fact is, however, that these various factors are often perceived as different in intensity, so that they can be ranked by strength, from the least to the most preferred – forming what Optimality Theory calls *constraint rankings* (Prince & Smolensky, 1993). In music perception, grouping constraints, especially those originating from Lerdahl and Jackendoff's local Grouping Preference Rule 3 (GPR3, GTTM), have been studied a number of times so far (Deliege, 1987; Clarke & Krumhansl, 1990; van der Werf & Hendriks, 2004; Frankland & Cohen, 2004). However, even though metrical segmentation has been one of the most widely studied aspects of music perception (e.g. Rothstein, 1989; Parncutt, 1994; Roberts, 1996; Hasty, 1997) and the venue of promising new theories of music cognition (along with GTTM, at least also Temperley, 2000; 2004;

Lerdahl, 2001), to our knowledge, there has still not been a true empirical investigation inducing individuals to construct “metrical Gestalten” on the basis of the preference rules suggested in GTTM.

We have therefore designed an empirical study with a triple goal: (a) to test whether metrical preference rules/constraints from the group MPR5 presented in GTTM indeed influence parsing choices, and, if yes, with what intensity; (b) to determine whether there are substantial differences in the perception of such metrical structures between musicians and nonmusicians; and (c) to find out whether there are differences in the perception of these structures if they are played in such a way as to comply with or contradict the parsers’ expectancies.

Hypotheses

The hypotheses we have put forward are in line with the three research goals defined above.

Test Hypothesis

Metrical constraints as proposed in MPR5 from GTTM can be ranked by strength. The ranking remains relatively stable regardless of different musical experience (musicians/nonmusicians) or disrupted expectancies (expected/unexpected ordering of beats).

Auxiliary Hypotheses

1. Musicians and nonmusicians have equal internal constraint rankings, relative to their success in the segmentation task. In other words, even if musicians have more correct responses, the internal order of constraints will be similar in the two groups.

2. When their expectancies are disrupted, participants respond by decreased number of correct responses and increased latencies. However, the overall constraint ranking remains relatively stable again.

Procedure

The central question we wished to address in the research was the justifiability of the concept of preference rules. For this reason, along with the guidelines offered in MPR5 from GTTM, we constructed 15 metrical stimuli, comparable by virtually all properties but the targeted constraint. All stimuli were so devised as to repeat the targeted metrical pattern ten times in a row (through ten measures). The individuals were asked to segment each stimulus where they felt they should do so, by pressing a button when certain they had heard a stressed beat in the sequence, in any measure.

The sample comprised 120 randomly selected undergraduate students of the University of Nis, Serbia (*N*=120, *m*=60, *f*=60, *mean age* 21.06, *STD* 1.57, *range* 18-25). They were classified into four strata by education, as follows: 30 students of music, 30 students of social sciences and humanities, 30 students of natural sciences and mathematics, and 30 students of IT and engineering sciences, reflecting the general organizational structure of the University. For the purposes of this paper, we shall only discuss the results of musicians (*n*=30) and nonmusicians (*n*=90), where a musician is defined as a person receiving university-level music education.

The perception task was carried out individually. The respondents were explained that they were about to hear metrical patterns, where ‘perception of rhythm’ would be examined, and that there were no correct or incorrect responses. Musicians were additionally asked to respond by their initial feeling, and to exclude their musical education as much as possible while carrying out the task. The stimuli were played on a laptop computer with a pair of headphones for the participants, where their task was to press the spacebar on the laptop *only once*, upon hearing what they believed was the stressed beat. Prior to this, the participants had been played a simple example, a 100bpm 4/4 beat repeated 10 times with the first of four tones played in *forte* dynamics. This was done as we wanted to be sure that they understood the meaning of ‘stressed’ and could practice pressing the button.

The software for data presentation was made specifically for this purpose by a professor of the local IT school. The experimenter had full control of the software (stopping the stimulus and the program, repetition, turning the volume up or down). The task required interaction between the participants and the computer, as they were expected to press the space bar upon hearing a stressed beat. The pressure was registered by the software, where the time that elapsed from the inception of the targeted stressed beat in the particular measure to the moment of pressing was recorded in a separate log file. The laptop was set up in such a way as to reduce possible undesired software latencies to a minimum (it contained only the essentials of the operating system Windows XP and the experimental software, whose priority was manually set to maximum by the programmer).

Based on suggestions from GTTM, the stimuli were made on a personal computer, with the help of sequencing and sound processing software. Samples from the standard 128-sample set of MIDI instruments were used. The sequences were played by the sample simulating the grand piano.

GTTM proposes six variants of the MPR5 rule:

> Prefer a metrical structure in which a relatively strong beat occurs at the inception of either: a. a relatively long pitch-event, b. a relatively long

> duration of a dynamic, c. a relatively long slur, d. a relatively long pattern of articulation, e. a relatively long duration of a pitch in the relevant levels of the time-span reduction, or f. a relatively long duration of a harmony in the relevant levels of the time-span reduction (harmonic rhythm). (Lerdahl & Jackendoff, 1983: 84)

For this reason we made twelve stimuli: six following the above suggestions from GTTM, starting with a stressed beat (complying with expectancies) and six corresponding stimuli, equal to the former six in all properties, except for their beginning – as they did not start with the first, stressed, suggested beat, but with another beat from the measure (relatively unstressed, disrupting expectancies). There were additional three "fake" stimuli: they had nothing to do with GTTM metrical preference rules, but were used to distract the participants' attention and prevent them from improving their result towards the end of the task by learning. Another precaution in that respect was the software randomization of the order of the 15 stimuli.

Each metrical sequence contained ten measures, and the examples below present the transcriptions of the six stimuli pairs. Stimuli to the left are "expected" (complying with expectancies) and those to the right are "unexpected" (disrupting expectancies). The position of the targeted stressed beat, i.e. constraint, is marked with an asterisk (*). We hope that the examples show that the stimuli were equal in all respects but the targeted constraints – as much as possible. We purposefully did not produce identical stimuli for reasons of monotony, fear of the learning effect, and the need for them to comply as much as possible with the GTTM originals. However, they were all played on the same instrument, in the same key (C major), with the same articulation, dynamic, and tempo, except when one of these musical elements was to be the suggestive factor. All examples but one had a 4/4 beat. Pitch changes were also as steady as possible, without any sudden tonal leaps, while all melodic lines clustered around the middle C. As examples were composed by the author, they could fit more easily with the proposals from GTTM, but the question of experimenter bias still remains open. Yet, if the stimuli are taken from the musical literature, one confronts the problem of participants' potential familiarity with the examples. The complexity of such music is often also an issue. Hence, we do not believe one method should *per se* be favored against another. Examples below provide the first two measures of each stimuli pair:

Figure 1

As mentioned above, the software calculated the response latency from the occurrence of the stressed beat in any measure in which the particular subject pressed the button. It marked as correct any response which occurred at most 50ms *before* and 250ms *after* the sounding of the stressed beat in any measure. This criterion is of course somewhat arbitrary, as it poses an artificial window for accepting some responses as correct. However, the decision was not fully random: in none of the metrical examples was the time that elapsed from the principal stressed beat to the adjacent relatively unstressed beat shorter than 300ms. By allowing for the 250ms latency, we were thus rather benevolent to our participants, as we labeled as correct any response occurring after the stressed beat, and immediately before the relatively unstressed beat that followed. Going further than this would have made no sense, as any larger latency would have bordered on the incorrect zone. As for the 50ms prior to the sounding of the stressed beat, it was a “rush” that we allowed for we feared that some participants, especially musicians, might have strong expectancies and press the space bar a bit earlier than the occurrence of the note itself. Labeling the responses of such “quick thinking” participants as incorrect could have been unfair. Thus, we ended up with a third-second range for each stressed beat, which was, we hope, quite enough to prevent even the slowest or most cautious participants from making an accidental wrong choice. Students who claimed they had made an accidental press were not allowed to retake the task for that stimulus. Those who failed to press the button within the ten measures in the sequence were not allowed to repeat the task either.

This research design helped us obtain three types of metrical variables. Based on the latency range described above, the software first tested whether the participant had at all reacted to the suggested stressed beat. If not, this was an immediate incorrect response, where further calculation stopped. These data helped us determine the frequencies and percentage of correct responses to all stimuli, providing us with preliminary rankings of constraints. For those students who did guess the location of the stressed beats correctly, the software calculated the measure in which the response occurred, and also the response latency in milliseconds from the moment of the stress. Along with the data on correct responses, these two additional pieces of information allowed us to discuss what changes occurred in the perception of our metrical examples in case of deliberately disrupted expectancies.

Results

Constraint Rankings – Entire Population

Table 1 presents a comparative overview of correct and incorrect responses to the six stimuli pairs (expected suggestion to the left, unexpected suggestion to the right), for the entire sample (*N*=120). The results of the chi-square test for each pair are also provided below, denoting the probability that the different distribution of two responses was not accidental – i.e. that our playing with the participants' expectancies did cause significant changes in the segmentation. These simple statistical tests were not relevant to the constraint rankings that emerged from the results, but were later used as one of three elements to decide whether the altered ordering of the suggestive stressed beats in the stimuli indeed disrupted the participants' expectancies (along with the average latency and number of measure in which the reaction occurred, see 4.3).

Table 1

The numbers and percentages of correct responses (to the expected suggestion, unexpected suggestion, and totals) were then used to create a preliminary ranking of constraints for the entire sample, as provided in Table 2.

Table 2

Except for the constraint 'length' (MPR5a - fourth position in unexpected suggestions, and fifth position in expected suggestions), the ordering of constraints is identical. The same goes for the totals, provided to the right.

The ranking of preference rules/constraints from the group MPR5 on the sample was thus as follows:

dynamic > harmony >> slur > pitch > length >> articulation

We have taken over the notation of Optimality Theory (Prince & Smolensky, 1993) where '>' marks a difference in intensity, and '>>' means a pronounced difference in intensity. In our case differences in the frequency of responses between adjacent constraints were not sufficient to justify a generalization. It may be seen, though, that differences between three *groups* of constraints, as bracketed below, showed to be significant ($p < .05$, equality of proportions test, see Appendix A). Therefore, ‘>’ marks the difference on the sample, and should be used as illustration only. On the other hand, ‘>>’ marks the difference in the population, and it seems to be relevant.

(dynamic > harmony) >> (slur > pitch > length) >> (articulation)

Somewhat tentatively, we propose that the three groups be labeled “physical stress” (as dynamic and harmonic changes produce significant change to the sound produced), “melodic stress” (as slur, pitch, and length changes all have to do with the change in the melodic line of the tune), and “ornamental stress” (as type of articulation in music is often considered an ornamental device). Thus, with our sample size, we get the following preliminary ranking of three constraint groups, as generalized from GTTM MPR5:

physical stress >> melodic stress >> ornamental stress

Constraint Rankings – Musicians vs. Nonmusicians

Table 3 provides distributions of responses to the six stimuli pairs (complying and not complying with expectancies), divided by musicians and nonmusicians (n_1=30, n_2=90). Below each stimuli pair, chi-square results are provided: this time, the test calculated the statistical significance for the difference in the response distribution (to both expected and unexpected suggestion stimuli) between musicians and nonmusicians. Again, this calculation did not directly influence the constraint rankings for musicians and nonmusicians, but it helped us test how much, irrespective of internal ranking, musicians differed from nonmusicians in the absolute numbers of correct responses. These results show that in 7 out of 12 stimuli, musicians responded more accurately than nonmusicians, as our auxiliary hypothesis 1 had anticipated ($p < .05$). In 5 stimuli, however, there was no statistically significant difference between the success of musicians and nonmusicians in the segmentation task. This especially applies to unexpected suggestions, where musicians and nonmusicians seem to have been equally confused after their expectancies had been disrupted (two different distributions out of six stimuli pairs).

Table 3

Table 4 provides the ranking of constraints (expected suggestion, unexpected suggestion, and totals) in musicians and nonmusicians.

Table 4

Once again, the constraint “length” occupies the fourth position in musicians, and the fifth position in nonmusicians. Looking at the totals (the third column) we find the ranking of individual constraints provided below. The equality of proportions test has helped us again classify the constraints for the two strata into three macro-groups (see Appendix B). However, with musicians, the calculation allows us to claim that slur is different from length, but not from harmony or dynamic in the population. With nonmusicians, pitch can be said to be different from length, but not from dynamic, harmony or slur ($p < .05$):

Musicians: (harmony > dynamic>slur)>> (length > pitch)>>articulation

Nonmusicians: (dynamic > harmony>slur>pitch)>>(length)>>articulation

Therefore, divided by musicians and nonmusicians, we suggest some alterations to the generalized three-group model offered above:

Musicians: physical stress + slur >> melodic stress >> ornamental stress

Nonmusicians: physical stress + slur + pitch >> melodic stress >> ornamental stress

The classification into three groups remains. We thus suggest that, regardless of their relatively different achievement in absolute figures, the internal constraint rankings of musicians and nonmusicians from our population are similar. Further research should test on a larger sample whether they might even be identical.

Expectancy Revisited

The final segment of the study discusses the well-known issue of expectancy (as tested recently in music perception at least by Large & Palmer, 2002; Jongsma, Quiroga & VanRijn, 2004; in language perception by Quene & Port, 2005). We anticipated that starting the sequence with an unstressed beat, which disrupted the 'natural', 'logical' sequencing, would result in fewer correct responses, reaction in more distant measures, and prolonged response times in any given measure. The data for the difference in distribution of responses to the expected and unexpected sequences are given through chi-square tests in Table 1: they show that, in the entire sample, in all stimuli pairs but two (MPR5a, d: length, articulation), the ratio of correct and incorrect responses significantly differs in expected and unexpected stimuli pairs. Stressing the same point from a different angle, Table 5 presents average latencies in milliseconds to the expected and unexpected suggestion stimuli from the pair (calculated from the inception of the measure, only for those participants who correctly guessed the stress in *both* stimuli), followed by 95% confidence interval calculations. Except for the first stimulus pair (constraint MPR5a: length), the remaining five stimuli show a statistically significant latency increase in sequences with unexpected suggestions.

Table 5

In short, the reduced number of hits (with statistical significance on the level $p < .05$ except for length and articulation), and prolonged average latencies (in all pairs but length, CI 95%) testify, once again, to expectancy being a relevant phenomenon in metrical perception. Not much could be seen from the particular measure in which the hit was made, as participants generally pressed the spacebar in the third, fourth or fifth measure, regardless of the correctness of their response (the mean measure in which the hit occurred ranged from 2.43 to 4.24). In other words, it seems to us that factors inducing them to press the button in a particular measure might have been partly extramusical.

There is one more result suggesting how important expectancies are: the dramatic drop of musicians' accuracy in the unexpected stimuli pairs, resulting in the fact that the statistical significance for the difference between musicians' and nonmusicians' achievement all but vanished in the unexpected stimuli group (except for MPR5f - harmony, Table 3, chi-square results to the right). Even trained musical professionals seem to have problems constructing metrical Gestalten when their expectations are deliberately disrupted.

In terms of the constraint rankings classified by expected and unexpected stimuli, the result follows (equality of proportions test, $p< .05$, see Appendix C):

Expected: (harmony > dynamic)>> (slur > pitch)>>(length> articulation)

Unexpected: (harmony> dynamic)>>(slur>pitch>length)>>(articulation)

Or, using our umbrella classification:

Expected: physical stress >> melodic stress >> ornamental stress + length

Unexpected: physical stress >> melodic stress >> ornamental stress

Expectancies are an issue that has to be considered in any investigation of metrical perception. In our study, their influence, especially on musicians, was obvious. Still, there was little difference between the internal ordering of constraints in the expected and unexpected stimuli group (except for the position of length, MPR5a).

In short, the preliminary conclusion appears valid stating that, with small exceptions, in our population metrical constraint ranking seemed to be a relatively stable phenomenon (not strongly correlated with either musical education or disrupted expectancies). The result should be fine-tuned in further studies.

Discussion

Viewing the results in light of our hypotheses, we suggest that the metrical preference rules from the group MPR5 proposed in GTTM may have some empirical validity. The constraints indeed appeared to differ in intensity, according to our test hypothesis. The exact ranking, however, remains unresolved, as our sample size and stimulus design failed to account for the position of adjacent constraints in the entire population. We still managed to make statistically valid generalizations for three groups of constraints that we labeled “physical”, “melodic”, and “ornamental” stress factors (Appendix A).

As for “physical stress”, the change in dynamics and the introduction of the harmonic triad in the lower voices showed to be the strongest segmentation factors in this study. All else being equal, the physically stronger element will become cognitively, and thus structurally, more relevant. This was only to be expected, especially with nonmusicians. The importance of the harmonic background for the inference of stressed beats has not surprised us, either: although a higher-order musical factor, chord sequencing seems to be so important for occidental ears that both musicians and nonmusicians consider this suggestion very relevant for determining meter, especially if it is well-formed, as was the case in our example (plagal cadence I-IV-I). In our research, this was the only stimulus pair in which we directly confronted two constraints (length and harmony). Even if it is true that these are “different order” preference rules, and cannot be compared so easily, in our examples it turned out that harmony was the definite winner. In Gestalt psychology terms, what we confronted here was “proximity” and “figure background”, where the latter seems to have clear advantage in metrical perception, an issue that might be given some consideration in further research.

The “melodic stress” group ranked second, and it consisted of three individual constraints: slur, pitch, and length. If a stronger note and prominent harmony, that appeared in the first group, are partly differentiated from the melodic line and provide strong impetus to the parser to segment the musical structure in the exact location, with slur and pitch change, there is no such “additional” effect. The parser rather concentrates on the melodic progression and must infer meter during this process. Our slur and pitch examples (MPR5 c, e: Figure 1) indeed urged the participants to focus on the pitch progressions, where there was nothing else to rely on while inferring meter, so that the task was definitely more difficult. Length, on the other hand, contained only two notes identical in all features but duration (MPR 5a: Figure 1). This melodic line was even simpler and there were even fewer elements for the participants to count on while deciding on the stressed beat, which is seen in the constraint ranking (esp. Table 4). Subsequent discussion with some musicians revealed to us that they gave this example a lot of thought before deciding. For some, the longer tone was stressed, for others, this was the shorter tone. In other words, it seems that the musically trained participants *perceived* our desired constraint here, but failed to agree with us on the *interpretation* of its importance. Thus, sheer duration of tones, in the absence of any other suggestion, cannot really be taken as a strong predictive factor for metrical segmentation.

Articulation was the last constraint in the ranking in all calculations, significantly weaker in intensity than its preceding constraints. We are partially to be held responsible for this result, as the musical example that we offered was indeed a bit more difficult, albeit almost exactly copied from GTTM p. 82, ex. 4.29 (succession of two sixteen-note quadruplets and triplets in a 4/4 beat, at 100bpm, MPR5d, Figure 1). Yet, although we do accept some blame for the weak result, due to the complexity of the stimulus and a slightly faster tempo, caused by the introduction of the sixteenth notes, we fear that the stimulus was insufficiently discriminative also due to the nature of the suggestion. For our participants, the triplets were equally possible bearers of the stress as were the quadruplets, and this, as we labeled it, “ornamental” factor did not have any significant predictive value, so it ended up last in all constraint rankings.

As it may be, the organization of six constraints into three more general groups seems to hold, and we hope that further research will fine-tune this result.

Our two auxiliary hypotheses have been partly corroborated. In terms of hypothesis 1, musicians did have better results than nonmusicians in seven examples out of twelve ($p < .05$). Yet in five stimuli, four of which were in the unexpected suggestion group, there was no statistical significance for the different distribution of correct and incorrect responses. In other words, musicians were indeed much better when expectancies were left alone, but not particularly better when expectancies were disrupted. Whether this had to do with their lack of concentration while performing the task, or with the strong influence of expectancies as a limiting factor, as suggested in the literature, remains to be further studied. It is obvious, however, that musicians’ success significantly decreased in the unexpected suggestion group and that disrupted expectancies had something to do with such an outcome.

If classified as three macro-groups that we proposed above, the constraint rankings of musicians and nonmusicians were similar. Some caution is warranted here. If we attempt a generalization into three groups by strength (Appendix B), the calculation claims that ‘pitch’ (MPR5d) belongs to the first group in nonmusicians, and to the second group in musicians. This should be further tested, as it may, but need not, be a consequence of the fact the group of musicians had fewer participants (30 : 90). The remaining five constraints are equally classified in the two groups, however. For this reason, we hope that our result is important. It may shed some new light on the question which segments of musical perception are more universal than others. Perhaps metrical segmentation, rather than grouping, is the principal domain in which GTTM succeeded in searching for musical universals (at least in terms of the difference between trained and untrained ears).

Finally, the tenets of the second auxiliary hypothesis seem to be true. In the entire population, the answer distributions differed in four stimuli pairs out of six, where the expected group had significantly more correct answers, and the response latency was significantly longer in five examples out of six. Both tendencies suggest a strong influence of expectancies on metrical perception, yet without major changes in the ranking of constraints (except for the position of MPR5a, length, Appendix C). In our research, the average measure in which the button was pressed (1-10) was not a relevant factor, either for the segmentation of metrical patterns or for the ranking of constraints, which may be contested in future studies.

Conclusion

In short, we hope that this study has shown that the metrical preference rules from Group 5 proposed in GTTM do have some empirical value. Our results suggest that constraints may be ranked, although their precise ordering remains to be determined. When doing metrical segmentation tasks, musicians and nonmusicians differ in many respects, but their internal constraint rankings, generalized into three macro-groups, are very much alike. Finally, in metrical perception, expectancy remains a formidable construct, but it does not significantly influence rankings, either.

Unanswered questions remain, as do suggestions for further research. Although simultaneous work of a number of constraints in any musical piece cannot be avoided by definition, in this study, we have not *deliberately* confronted constraints in the same examples (except for MPR5f). We still hope that our stimuli were well constructed along with the guidelines from GTTM to produce some provoking results. The ranking that we got is preliminary and should be further tested. Further research could also more deeply consider latencies and the number of measure in which the constraint was responded to as variables influencing the final ordering of constraints. We considered these data in relation to the problem of expectancy. In calculating the ranking, though, the only factor that we took into account was the sheer correctness of the response. Finally, we did not get the statistical significance for the ordering of all six constraints, but only of three broader groups. More precise ranking would require either a more sensitive construction of stimuli or a larger sample (or both).

Irrespective of these potential remarks, if anything, we are more confident now that preference rules/constraints should be favored over binary choices, at least in the segmentation of metrical patterns by western ears. This would in itself be a remarkable prediction of the often praised, but also criticized, quarter of a century old theory of music cognition.

References

Clarke, E. & Krumhansl, C. (1990). Perceiving musical time. *Music Perception, 7*, 213-251.

Deliege, I. (1987). Grouping conditions in listening to music: an approach to Lerdahl and Jackendoff's grouping preference rules. *Music Perception, 4*, 325-360.

Frankland, B. W., & Cohen, J. A. (2004). Parsing of melody: quantification and testing of the local grouping rules of Lerdahl and Jackendoff's A Generative Theory of Tonal Music. *Music Perception*, *21, 4*, 499-543.

Gilbers, D. & Schreuder, M. (2002). Language and music in optimality theory. *Rutgers Optimality Archive*, ROA 571-0103, http://roa.rutdgers.edu.

Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics*, vol. 3, (pp. 41-58), New York: Academic.

Hasty, C. (1997). *Meter as Rhythm*, Oxford University Press

Jackendoff, R. (2002). *Foundations of Language*, Oxford University Press

Jongsma, M., Quiroga, R., & van Rijn, C. (2004). Rhythmic training decreases latency jitter of omission evoked potentials in humans. *Neuroscience Letters 355*, 189-192.

Large, E. & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science Journal*, *26, 11*, 1-37.

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

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

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception 11*, 409-464.

Prince A. & Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University Center for Cognitive Science Technical Report 2.

Quene, H. & Port, R. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica*, *62*, 1-13.

Roberts, S. (1996). *Interpreting Rhythmic Structures Using Artificial Neural Networks.* Unpublished dissertation, University of Wales, College of Cardiff

Ross, J.R. (1970). *Constraints on Variables in Syntax.* Unpublished dissertation, Massachusetts Institute of Technology

Rothstein, W. (1989). *Phrase Rhythm in Tonal Music.* New York: Schirmer Books

Temperley, D. (2000). Meter and grouping in African music: a view from music theory. *Ethnomusicology*, *44, 1*, 65-96.

Temperley, D. (2004). *The Cognition of Basic Musical Structures.* MIT Press

van der Werf, S. & Hendriks, P. (2004). A constraint-based approach to grouping in language and music. In: R. Parncutt et al. (Eds.) *Proceedings, Conference on Interdisciplinary Musicology*, *CIM04*, Graz, Austria.

Wertheimer, M. (1923). Laws of organization in perceptual forms. in W.D. Ellis (Ed.). *A Source Book of Gestalt Psychology* (pp. 71-88). Routledge.

Appendix A

Equality of Proportions Probabilities (Entire Sample)

Constraint	Dynamic	Harmony	Slur	Pitch	Length	Articulation
Dynamic	*					
Harmony	0.0467	*				
Slur	0.000	0.0230	*			
Pitch	0.000	0.0002	0.1454	*		
Length	0.000	0.0000	0.0181	0.4463	*	
Articulation	0.000	0.0000	0.0000	0.0001	0.0020	*

Population**:** dynamic, harmony >> slur, pitch, length >> articulation ($p < .05$)

Appendix B

Equality of Proportions Probabilities. Musicians and Nonmusicians

Constraint	Harmony	Dynamic	Slur	Pitch	Length	Articulation
Harmony	*	0.0163	0.0000	0.0000	0.0000	0.0000
Dynamic	0.8406	*	0.0686	0.0074	0.0001	0.0000
Slur	0.1840	0.2581	*	0.3821	0.0431	0.0000
Pitch	0.0065	0.0113	0.1476	*	0.2471	0.0005
Length	0.0171	0.0282	0.2745	0.7178	*	0.0166
Articulation	0.0000	0.0001	0.0023	0.0923	0.0422	*

Musicians, below the diagonal / nonmusicians, above the diagonal

Musicians: dynamic, harmony, slur >> pitch, length >> articulation ($p < .05$)

Nonmusicians: dynamic, harmony, slur, pitch >> length >> articulation ($p < .05$)

Appendix C

Equality of Proportions Probabilities. Expected and Unexpected Suggestions

Constraint	Dynamic	Harmony	Slur	Pitch	Length	Articulation
Dynamic	*	0.043	0.0196	0.0004	0.0139	0.0000
Harmony	0.495	*	0.747	0.1168	0.6510	0.0002
Slur	0.0003	0.0027	*	0.2113	0.8965	0.0007
Pitch	0.0000	0.0001	0.3527	*	0.2623	0.0232
Length	0.0000	0.0000	0.0030	0.0330	*	0.0013
Articulation	0.0000	0.0000	0.0000	0.0014	0.2534	*

Expected, below the diagonal / unexpected, above the diagonal

Expected: dynamic, harmony >> slur, pitch >> length, articulation ($p < .05$)

Unexpected: dynamic, harmony >> slur, pitch, length >> articulation ($p < .05$)

Author Note

Mihailo Antovic, Faculty of Philosophy, University of Nis, Serbia.

This article is partially based on the dissertation entitled *Optimality and Metaphor Theory in Music and Language Cognition*, defended in November 2007 at the Faculty of Philosophy, University of Nis, Serbia. Some of the preliminary findings were reported as a poster presentation at the conference *Music and Language as Cognitive Systems*, University of Cambridge, UK, May 2007.

I wish to express gratitude to the dissertation supervisor, Prof. Djordje Vidanovic, committee members Prof. Aleksandar Kostic, Prof. Biljana Misic Ilic and Prof. Jadranka Hofman Jablan. I would also like to thank Nenad Popovic, M.A. for his invaluable help with data analysis and statistics. Milan Savic, Ph.D., programmed the stimulus presentation software, and Martin Jovanovic, Lidija Ristic, and Milica Tasic assisted me in conducting the experiments. Without their help, this paper would never have been written. The responsibility for any errors or inconsistencies remains exclusively my own.

Address for correspondence: Mihailo Antovic, Department of English, Cirila i Metodija 2, office 418, Faculty of Philosophy, University of Nis, 18000 Nis, Serbia, mantovic@gmail.com.

Table 1

Total Responses to Expected and Unexpected Suggestions - Entire Sample (N=120)

Constraint, correctness of response	*N (%)*	*N (%)*	*N (%)*
Constraint 1: Length (MPR5a)	Expected	Unexpected	Total
Incorrect	83 (69.2)	79 (65.8%)	162 (67.5)
Correct	37 (30.8)	41 (34.2%)	78 (32.5)
Total	120 (100)	120 (100)	240 (100)
Pearson χ^2 = 0.3039 df=1 p = 0.581			
Constraint 2: Dynamic (MPR5b)	Expected	Unexpected	Total
Incorrect	32 (26.7)	60 (50.0)	92 (38.3)
Correct	88 (73.3)	60 (50.0)	148 (61.7)
Total	120 (100)	120 (100)	240 (100)
Pearson χ^2 = 13.8190 df=1 p = 0.000			
Constraint 3: Slur (MPR5c)	Expected	Unexpected	Total
Incorrect	60 (50.0)	78 (65.0)	138 (57.5)
Correct	60 (50.0)	42 (35.0)	102 (42.5)
Total	120 (100)	120 (100)	240 (100)
Pearson χ^2 = 5.5243 df=1 p = 0.019			
Constraint 4: Articulation (MPR5d)	Expected	Unexpected	Total
Incorrect	91 (75.8)	101 (84.2)	192 (80.0)
Correct	29 (24.2)	19 (15.8)	48 (20.0)
Total	120 (100)	120 (100)	240 (100)
Pearson χ^2 = 2.6042 df=1 p = 0.107			

Constraint, correctness of response	*N(%)*	*N(%)*	*N(%)*
Constraint 5: Pitch (MPR5e)	Expected	Unexpected	Total
Incorrect	67 (55.8)	87 (72.5)	154 (64.2)
Correct	53 (44.2)	33 (27.5)	86 (35.8)
Total	120 (100)	120 (100)	240 (100)
Pearson χ^2 = 7.2486 df=1 p = 0.007			
Constraint 6: Harmony (MPR5f)	Expected	Unexpected	Total
Incorrect	37 (30.8)	76 (63.3)	113 (47.1)
Correct	83 (69.2)	44 (36.7)	127 (52.9)
Total	120 (100)	120 (100)	240 (100)
Pearson χ^2 = 25.4366 df=1 p = 0.000			

Table 2

Ranking of Metrical Constraints. Correct Responses. Entire Sample (N=120)

	Expected (120)		Unexpected (120)		Total (240)	
Constraint	*N*	*%*	*N*	*%*	*N*	*%*
Dynamic	88	73.3	60	50.0	148	61.7
Harmony	83	69.2	44	36.7	127	52.9
Slur	60	50.0	42	35.0	102	42.5
Pitch	53	44.2	33	27.5	86	35.8
Length	37	30.8	41	34.2	78	32.5
Articulation	29	24.2	19	15.8	48	20

Table 3.

Total responses. Musicians vs. Nonmusicians (n_1=30 vs. n_2=90).

Constraint 1: Length (MPR5a)							
	Expected Suggestion				Unexpected Suggestion		
	Musicians	Nonmusicians	Total		Musicians	Nonmusicians	Total
Incorrect	17	66	83	Incorrect	14	65	79
	56.7%	73.3%	69.2%		46.7%	72.2%	65.8%
Correct	13	24	37	Correct	16	25	41
	43.3%	26.7%	30.8%		53.3%	27.8%	34.2%
Total	30	90	120	Total	30	90	120
	100%	100%	100%		100%	100%	100%
Pearson χ^2=2.9306 df=1 p=0.087				Pearson χ^2=6.5329 df=1 p=0.011			

Constraint 2: Dynamic (MPR5b)							
	Expected Suggestion				Unexpected Suggestion		
	Musicians	Nonmusicians	Total		Musicians	Nonmusicians	Total
Incorrect	4	28	32	Incorrect	15	45	60
	13.3%	31.1%	26.7%		50.0%	50.0%	50.0%
Correct	26	62	88	Correct	15	45	60
	86.7%	68.9%	73.3%		50.0%	50.0%	50.0%
Total	30	90	120	Total	30	90	120
	100%	100%	100%		100%	100%	100%
Pearson χ^2=3.634 df=1 p=0.057				Pearson χ^2=0.000 df=1 p=1			

Constraint 3: Slur (MPR5c)

	Expected suggestion				Unexpected suggestion		
	Musicians	Nonmusicians	Total		Musicians	Nonmusicians	Total
Incorrect	8	52	60	Incorrect	17	61	78
	26.7%	57.8%	50.0%		56.7%	67.8%	65.0%
Correct	22	38	60	Correct	13	29	42
	73.3%	42.2%	50.0%		43.3%	32.2%	35.0%
Total	30	90	120	Total	30	90	120
	100%	100%	100%		100%	100%	100%

Pearson χ^2=8.7111 df=1 p=0.003 | Pearson χ^2=1.2210 df=1 p=0.269

Constraint 4: Articulation (MPR5d)

	Expected suggestion				Unexpected suggestion		
	Musicians	Nonmusicians	Total		Musicians	Nonmusicians	Total
Incorrect	17	74	91	Incorrect	25	76	101
	56.7%	82.2%	75.8%		83.3%	84.4%	84.2%
Correct	13	16	29	Correct	5	14	19
	43.3%	17.8%	24.2%		16.7%	15.6%	15.8%
Total	30	90	120	Total	30	90	120
	100%	100%	100%		100%	100%	100%

Pearson χ^2=8.0182 df=1 p=0.005 | Pearson χ^2=0.208 df=1 p=0.885

Constraint 5: Pitch (MPR5e)

	Expected suggestion				Unexpected suggestion		
	Musicians	Nonmusicians	Total		Musicians	Nonmusicians	Total
Incorrect	9	58	67	Incorrect	24	63	87
	30.0%	64.4%	55.8%		80.0%	70.0%	72.5%
Correct	21	32	53	Correct	6	27	33
	70.0%	35.6%	44.2%		20.0%	30.0%	27.5%
Total	30	90	120	Total	30	90	120
	100%	100%	100%		100%	100%	100%

Pearson χ^2=10.8251 df=1 p=0.001 Pearson χ^2=1.1285 df=1 p=0.288

Constraint 6: Harmony (MPR5f)

	Expected suggestion				Unexpected suggestion		
	Musicians	Nonmusicians	Total		Musicians	Nonmusicians	Total
Incorrect	3	34	37	Incorrect	14	62	76
	10.0%	37.8%	30.8%		46.7%	68.9%	63.3%
Correct	27	56	83	Correct	16	28	44
	90.0%	62.2%	69.1%		53.3%	31.1%	36.7%
Total	30	90 (100%)	120	Total	30	90 (100%)	120
	(100%)				(100%)		100%)

Pearson χ^2=8.1407 df=1 p=0.004 Pearson χ^2=4.7847 df=1 p=0.029

Table 4.

Ranking of Metrical Constraints. Correct Responses. Musicians vs. Nonmusicians. (n_1=30, n_2=90)

Musicians							Nonmusicians						
	Exp(30)		Unexp(30)		Total(60)			Exp(90)		Unexp(90)		Total(180)	
Constraint	*N*	%	*N*	%	*N*	%	Constraint	*N*	%	*N*	%	*N*	%
Harmony	27	90.0	16	53.3	42	70.0	Harmony	62	68.9	45	50.0	107	59.4
Dynamic	26	86.7	15	50.0	41	68.3	Dynamic	56	62.2	28	31.1	84	46.7
Slur	22	73.3	13	21.7	35	58.3	Slur	38	42.2	29	32.2	67	37.2
Pitch	21	70.0	6	20.0	27	45.0	Pitch	32	35.6	27	30.0	59	32.8
Length	13	43.3	16	53.3	29	48.3	Length	24	26.7	25	27.8	49	27.2
Articulat.	13	43.3	5	16.7	18	30.0	Articulat.	16	17.8	14	15.6	30	16.7

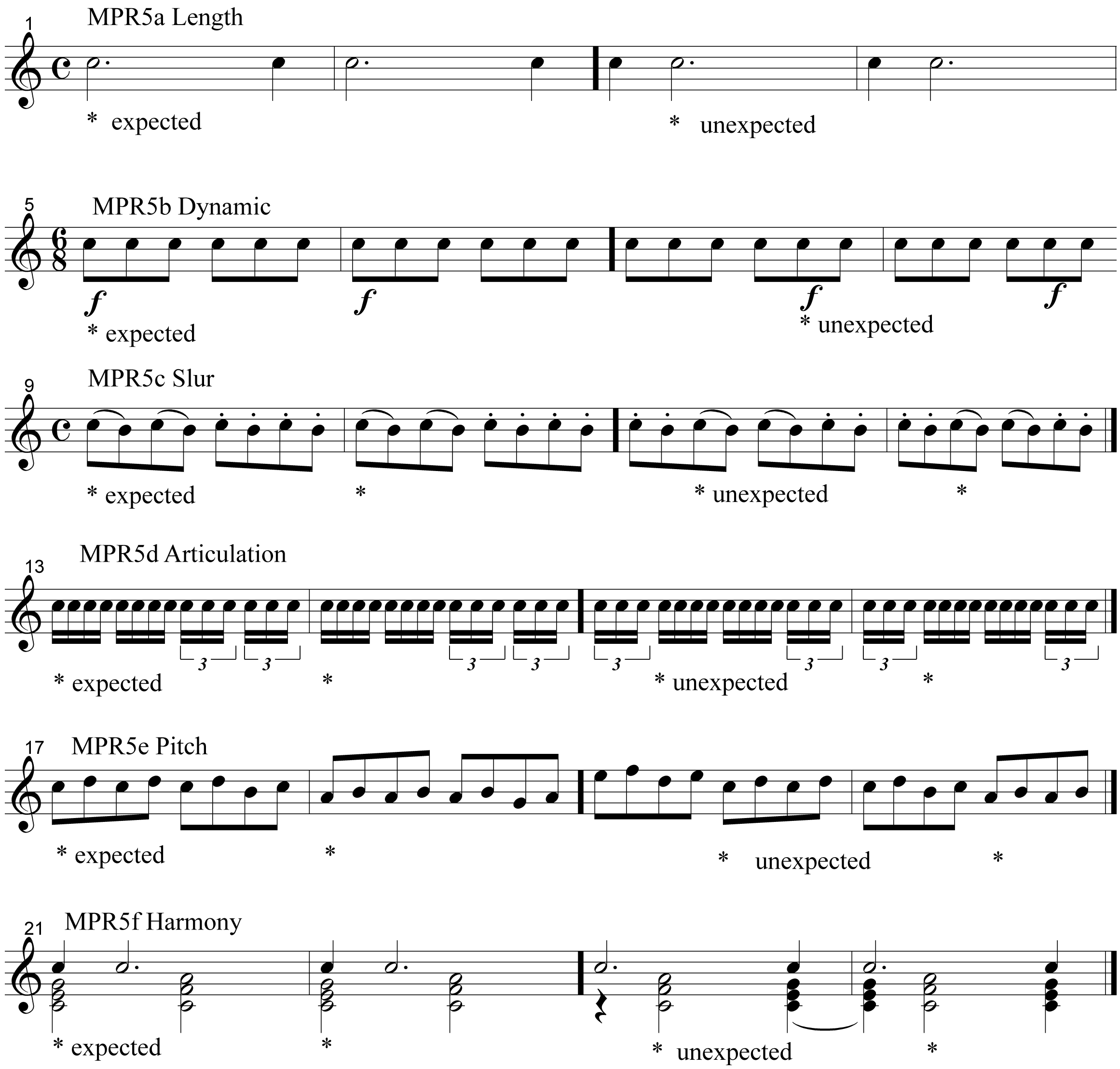
Table 5.

Expectancy. Average Response Latencies (in ms), Number of Participants who Correctly Located the Stress, Standard Deviation (Entire Population, N=120)

Expected Suggestion				Unexpected Suggestion			
Constraint	Avg.latency	*N*	*STD*	Constraint	Avg.latency	*N*	*STD*
1. Length	68.8919	37	128.63747	1. Length	68.9756	41	99.42597
	CI ± 3.65, p< .05				CI ± 2.54, p< .05		
2. Dynamic	70.0227	88	111.38150	2. Dynamic	122.3167	60	128.90628
	CI ± 2.23, p< .05				CI ± 2.87, p< .05		
3. Slur	76.8500	60	108.23716	3. Slur	62.1429	42	112.96893
	CI ± 2.63, p< .05				CI ± 3.21, p< .05		
4. Articul.	118.9655	29	153.45090	4. Articul.	182.5789	19	142.65783
	CI ± 4.51, p< .05				CI ± 5.37, p< .05		
5. Pitch	82.9245	53	113.22507	5. Pitch	96.6061	33	121.50950
	CI ± 2.86, p< .05				CI ± 3.76, p< .05		
6. Harmony	110.1566	83	111.08319	6. Harmony	116.9773	44	116.20581
	CI ± 2.26, p< .05				CI ± 3.18 p< .05		

Figure Captions

Figure 1. Stimuli pairs


1
MPR5a Length
* expected
* unexpected
5
MPR5b Dynamic
f
f
f
f
* expected
* unexpected
9
MPR5c Slur
* expected
*
* unexpected
*
13
MPR5d Articulation
3
* expected
*
* unexpected
*
17
MPR5e Pitch
* expected
*
* unexpected
*
21
MPR5f Harmony
* expected
*
* unexpected
*