FACTA UNIVERSITATIS
Series: **Linguistics and Literature** Vol. 13, N° 2, 2015, pp. 77 - 93

A LINGUISTIC CONSTRUCT INFORMS MUSICOLOGY: RANKING METRICAL CONSTRAINTS IN MUSIC PERCEPTION *

UDC 159.932:781.62

Mihailo Antović

Faculty of Philosophy, University of Niš, Serbia

Abstract. *The construct of perceptual constraints has become increasingly important in cognitive science in recent years, including the research at the intersection of linguistics and musicology. The present paper provides the results of an empirical study into the ordering of metrical preference rules/constraints from the group MPR5, as proposed in "A Generative Theory of Tonal Music" (GTTM, Lerdahl & Jackendoff, 1983). The theory predicts a preference for inferring strong beats on musical tones which exhibit a relatively prominent pitch change, dynamic, long slur, long pattern of articulation, long duration of a pitch in the relevant levels of a time-span reduction, and prominent harmony in the relevant level of a time-span reduction. A hundred and twenty randomly selected undergraduate students (30 musicians and 90 nonmusicians) were played twelve metrical sequences based on the examples of the rule MPR5 from GTTM, of which one half were constructed so as to comply with the participants' expectancies and another half so as to contradict them. The participants were prompted to press a button when certain they had heard a stressed beat. The distributions of responses suggest that the six constraints can be ranked into three larger groups, as follows: (dynamic, harmony), (pitch, slur, length), (articulation). Musicians achieved better results than nonmusicians, and the response latencies considerably rose in the stimuli contradicting expectancies, but the internal constraint rankings remained relatively stable irrespective of the two factors (musical training and inception of stimuli on the targeted beat). Given such results, metrical segmentation is hypothesized to be the principal contribution of GTTM which has stood the test of time.*

Key words: *musical meter, preference rules, constraints, optimality theory, generative theory of tonal music.*

Submitted September 10th 2015, accepted for publication November 12th, 2015
Corresponding author: Mihailo Antović
Faculty of Philosophy, University of Niš, Serbia
E-mail: mantovic@gmail.com
* The study was supported by the project no. 179013, conducted at the University of Niš – Faculty of Mechanical Engineering, and supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia. I also wish to express gratitude to Nenad Popović for assistance with the statistical analyses, Milan Savić for programming the stimulus presentation / reaction time measurement software, and Lidija Ristić, Milica Tasić and Martin Jovanović for helping me with the fieldwork.

1. INTRODUCTION

In psycholinguistics and much of cognitive science the final quarter of the twentieth century was marked by a 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 various 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; recently Gilbers & Schreuder, 2002; Jackendoff, 2002). Regardless of the name, the construct implies that temporal structures are parsed based on a set of physical changes in the quality of the stimulus, which are then perceived as clues as to how to organize the entire complex structure into meaningful wholes. The fact is, however, that these various factors are typically 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). Examples of such parsing procedures are numerous in language, especially in the domain of suprasegmental phonology, although optimality theorists have also offered some interesting insights on morphological, syntactic and (occasionally) semantic levels. A good test case from the Serbian language would be that of morphophonemic changes. We often witness indecision among native speakers on "which form is correct" in nouns such as *ćevabdžinica* and *buregdžinica* (shops where small minced meat balls and phyllo dough pie are sold). The suffix *-džinica* requires that the final consonant in the stem should become voiced (*burek → bureg; ćevap → ćevab*). However, this phonetically motivated prescriptive rule is often overlooked by laypersons, as they prefer the semantic criterion to the phonological one – what they eat is called *ćevap / burek* and not **ćevab* / **bureg*, and therefore their preferred final noun to denote the shop needs to be **ćevapdžinica* or **burekdžinica*. This is of course a violation of the prescribed norm, which is often dismissed as "uneducated", a phenomenon I gladly leave to sociolinguists to discuss. The fact remains, however, that "mistakes" are made by native speakers only when there is some kind of cognitive dissonance underlying the production of the final form. In this case, a semantic and a phonetic rule come into direct conflict, and one needs to prevail over the other in order for the final lexical item to be produced. If the phonetic criterion wins, we get the "prescribed" variant; if the semantic one prevails, we get the "mistake". Such conflicts, which may involve not only two but sometimes dozens of constraints, represent the core interest of Optimality Theory: indeed, its claim that "constraint violability" lies at the basis of universal grammar has caused considerable interest in generative linguistics, but also in the cognitive sciences at large since the 1990s.

Constraint-based theories come from strongly computationalist frameworks and they can in principle be applied to numerous cognitive phenomena. Thus they are not limited to studies of the language faculty. In fact, *The Generative Theory of Tonal Music* (GTTM, Lerdahl & Jackendoff, 1983), one of the most influential books to date on music cognition – importantly, inspired by Chomskyan generative linguistics – has earned much of its fame by introducing an approach to parsing musical structures based on constraint interactions.

In essence, GTTM is a formal, reductionist theory of music perception in the Western classical idiom. The approach it offers is metalinguistic since its principal epistemological assumptions derive from generative linguistics proper. Just like the linguistic "universal grammar", which is no longer a set of rules for "proper" speech but rather a descriptive theory targeting the native speaker's ability to form acceptable sentences in his or her mother tongue, the musical grammar in GTTM focuses on the "native listener's" ability to perceive a musical piece as conformant (or not) to his or her native "musical idiom", based on a series of deep, perhaps partly inborn intuitions. In the same reductionist manner in which the "surface structure" of a linguistic sentence is reduced to a set of abstract, "underlying" relations, the intuitions which the native listener has about the music he or she hears reduce the musical signal to a series of more abstract constituents. Typically, these inferred constituents are both (1) hierarchical in nature, where smaller segments compositionally form larger ones, and (2) expressive of structural dominance, where the listener intuitively perceives the salient elements in the structure on a variety of levels (metrical, melodic,…). The theory, in turn, attempts to predict the location of those dominant spots in pieces of real music, in the hope that some of those predictions would hold in actual experimental work with human participants.

GTTM targets musical perceptual hierarchies on levels it calls "grouping structure", "metrical structure", the structure of "time-span reductions", and the structure of "prolongational reductions". The present paper focuses on metrical relations, i.e. the inference of "patterns" from perceived successions of stressed and unstressed beats in a temporal auditory sequence. Like the other structures of GTTM, the metrical system is hierarchical, where "lower" metrical levels (e.g. a sequence of two beats, one of which is stressed) function as constituents in "higher" ones (e.g. four beats in a measure, with two relatively stressed beats, only one of which is, however, the most stressed of them all). In addition to the "well-formedness" rules, which are necessary for a temporal structure to be perceived as metrical at all, GTTM also introduces "preference rules", in which salient, stressed beats are inferred based on a number of clues present in the sound stimulus. Importantly, these clues can vary in strength, and thus form an internal "preferential hierarchy".

Therefore, the main question regarding metrical preference rules in GTTM could be formulated as follows: all else being equal, what is the preferred clue in the sound stimulus which induces listeners familiar with the musical idiom to mark a particular location in the musical flow as the location of metrical accent? Constraints on building musical melodic groups have been studied several times (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, Jackendoff, 2009, Rohrmeier & Koelsch, 2012, Hamanaka et al, 2013) and the venue of promising new theories of music cognition (along with GTTM, at least also Temperley, 2000; 2004; Lerdahl, 2001, London, 2012), to my knowledge, there has still not been a true empirical investigation inducing subjects to construct "metrical Gestalten" on the basis of the preference rules suggested in GTTM.

The present study therefore focuses on one particular metrical rule from GTTM which introduces several conflicting factors in metrical inference (MPR5). It mostly deals with the concept of "length" (duration) and is defined as follows in the original text: "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)"[1] (Lerdahl & Jackendoff, 1983: 84).

An empirical study has therefore been designed here, with a triple goal: (a) to test whether metrical preference rules/constraints from the group MPR5 presented in GTTM result in stable parsing choices; (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.

2. METHOD

2.1. Hypotheses

While GTTM vouches for the importance of constraints on numerous perceptual levels, it does not provide any predictions on their ordering and explicitly leaves this for future empirical work to determine. Yet given the overall 'universalist' undertone of GTTM, I start from the assumption that the ordering of the constraints – *any* hierarchy that we get, that is – will remain stable under various conditions. I therefore define one central and two derived hypotheses:

1. The ranking of metrical constraints as proposed in MPR5 remains stable regardless of the participants' different musical background (musicians/nonmusicians) or the ordering of beats in the stimulus (beginning on the stressed or unstressed beat).

1a. 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 overall, the internal ordering of the constraints remains stable in the two groups.

1b. When their expectancies are not fulfilled (i.e. when the stimuli do not begin on the targeted beat), participants respond by a decreased number of correct responses and increased response latencies. This is more prominent among nonmusicians. However, the overall constraint ranking remains relatively stable.

2.2. Procedure

The central question addressed in the research was the justifiability of the concept of preference rules. For this reason, along with the guidelines offered in MPR5 from GTTM (p.84), 15 metrical stimuli were constructed, comparable by numerous musical properties, but different in terms of the targeted constraint. All stimuli were so devised as to repeat the metrical pattern ten times in a row (through ten measures). The participants were asked to parse 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. As a rule, we did not repeat the sequence more than once, as exposure to ten successive instances of the

[1] Time-span reductions represent a higher level of musical cognitive organization, where individual pitch events are grouped into structurally relevant wholes - realizing, for instance, the harmonic link between the dominant and tonic chord, which comprises a short musical motive. Importantly, there is a central element in each time span, a "head" similar to heads found in linguistic phrases. Thus GTTM presents time-spans in a tree notation, reminiscing the analyses of Chomsky's generative grammar (for more details, cf. Lerdahl & Jackendoff, 1983, chapter 6).

metrical pattern was enough for the majority of participants to decide on their preferred location of the stressed beat. Yet in the few cases in which the participants could still not do so, we agreed to repeat the sequence for the second time (in effect, this resulted in the total of twenty successive targeted metrical patterns, which was enough for all participants to make a choice). One should note, though, that all the patterns in the sequence were identical so that there was no logical "ending" of the entire stimulus (e.g. a ritardando). This may have made the task a bit difficult for the participants, and remains to be addressed in future studies.

The sample comprised 120 randomly selected undergraduate students of the University of Niš, 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, I 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 participants were explained that they were about to hear metrical patterns, where 'the 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 the task was to press the spacebar on the laptop *only once*, upon hearing what they believed was the stressed beat. Prior to this, the respondents had been played a simple example, a 100bpm 4/4 meter signature repeated 10 times with the first of four tones played in *forte* dynamics. This was done to make 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 from the local university Electronic Engineering Department. 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.

Based on the suggestions from GTTM MPR 5 (as quoted in the introduction section above), 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. To test the six variants of the rule defined by Lerdahl and Jackendoff, we made twelve stimuli: they all followed the above suggestions from GTTM, yet six of them started with a stressed beat (i.e. complied with the participants' expectancies) and six did not start from the first, stressed beat, but from another beat from the measure (a relatively unstressed one, and thus they did not fulfill the participants' expectancies). There were also three additional "fake" stimuli: they had nothing to do with GTTM metrical preference rules, but were used to distract the subjects' 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 in Figure 1 present the transcriptions of the six stimulus pairs (two measures each, for illustration purposes). Stimuli to the left started with the targeted stressed beat, and were thus "expected" (complying with expectancies), while those to the right began on an unstressed beat, one which was not targeted, and were thus "unexpected" (disrupting expectancies). The position of the targeted stressed beat, i.e. constraint, is marked with an asterisk (*).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 C_5. The double meter and C major key were used for two reasons: to follow the original examples from Lerdahl and Jackendoff (1983: 80-82) and also to ensure relative comparability of the stimuli with one another. Examples below provide the first two measures of each stimulus pair:

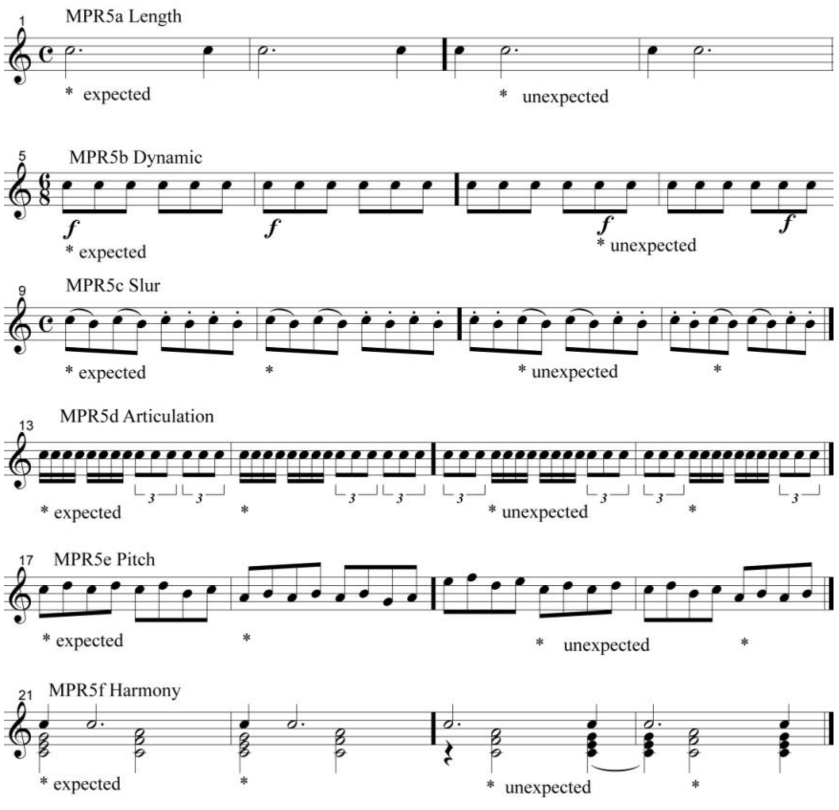

Fig. 1. Stimulus pairs

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 was based on the fact that 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 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 subjects, 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 "correct" range for each stressed beat – hopefully enough to prevent even the slowest or most cautious subjects 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, except in the very few situations in which they explicitly asked to do so, in which case the stimulus was repeated once.

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 opted for the suggested beat as stressed. 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 subjects 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 look into any changes in the perception of the metrical examples in case of expectancies that were deliberately not fulfilled.

3. RESULTS

3.1. Constraint rankings – the entire population

Table 1 presents a comparative overview of correct and incorrect responses to the six stimulus 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 the non-fulfillment of the participants' expectancies did cause significant changes in the segmentation.

It turned out there was a significant decrease in the number of correct responses when expectancies were not fulfilled in four stimuli out of six (all but *length* and *articulation*).

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			

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 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	20,0

Except for the constraint '*length*' (MPR5a – fourth position in unexpected suggestions, and fifth position in expected suggestions), the ordering of constraints is identical. The totals, provided to the right, reinforce the ranking from the 'expected' group.

The preference rules/constraints from the group MPR5 thus ranked in the following order:

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

I have taken over the notation of Optimality Theory (Prince & Smolensky, 1993) where '>' marks a difference in intensity, and '>>' denotes a pronounced difference in intensity. In the present case the differences in the frequency of responses between adjacent constraints were not sufficient to justify a statistically significant generalization. It may be seen, though, that the differences between three *groups* of constraints, as bracketed below, turned out to be statistically significant ($p< .05$, see Appendix A for the equality of proportions probabilities). Therefore, '>' marks a difference on the sample, and should be used as an illustration only. On the other hand, '>>' marks a difference in the population, which was statistically corroborated, and which represents the central finding of the present study.

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

3.2 Constraint rankings – musicians vs. nonmusicians

When one analyzes the distribution of "correct" and "incorrect" responses to the individual stimuli given by musicians and non-musicians, the following tendencies are conspicuous (for reasons of concision I do not provide the full tables here). It turns out that in 8 out of 12 stimuli musicians responded more accurately than nonmusicians, as our auxiliary hypothesis 1a had anticipated ($p< .05$). In 4 stimuli, however, there was no statistically significant difference between the success of musicians and nonmusicians in the segmentation task. The result becomes very interesting when one cross-links this with the type of stimulus: in the group with expected suggestions, musicians had a significantly larger number of correct responses than nonmusicians in *all* six stimuli; in the group with unexpected suggestions, however, musicians scored better than nonmusicians in only two examples out of six (*length* and *harmony*). The situation in which expectancies were not fulfilled was thus largely not intuitive to musicians either.

Table 3 provides the ranking of constraints (expected suggestion, unexpected suggestion, and totals) in musicians and nonmusicians.

Table 3 Ranking of metrical constraints. Correct responses. Musicians vs. nonmusicians. (n_1=30, n_2=90)

Musicians						
	Exp,30		Unexp,30		Total(60)	
Constraint	N	%	N	%	N	%
Harmony	27	90.0	16	53.3	42	70.0
Dynamic	26	86.7	15	50.0	41	68.3
Slur	22	73.3	13	21.7	35	58.3
Pitch	21	70.0	6	20.0	27	45.0
Length	13	43.3	16	53.3	29	48.3
Articulat.	13	43.3	5	16.7	18	30.0

Nonmusicians						
	Exp,90		Unexp,90		Total(180)	
Constraint	*N*	*%*	*N*	*%*	*N*	*%*
Harmony	62	68.9	45	50.0	107	59.4
Dynamic	56	62.2	28	31.1	84	46.7
Slur	38	42.2	29	32.2	67	37.2
Pitch	32	35.6	27	30.0	59	32.8
Length	24	26.7	25	27.8	49	27.2
Articulat.	16	17.8	14	15.6	30	16.7

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, also 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, in the entire population, the perception of targeted beats in the "slur" example was different from this perception in the "length" example, but not from the corresponding perception in examples labeled "harmony" or "dynamic". With nonmusicians, the "pitch" example can be said to have received a different number of "correct" metrical hits from the "length" example, but not from the stimuli labeled "dynamic", "harmony" or "slur" ($p< .05$):

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

The classification into three groups remains. We thus suggest that the internal constraint rankings of musicians and nonmusicians from our population are similar, but not quite identical. Further research should test this nuance on a larger sample.

3.3. Expectancies

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, Huron, 2006; in language perception by Quene & Port, 2005). The anticipation was that starting the sequence with an unstressed beat, which failed to fulfill the 'natural', 'logical' sequencing, would result in fewer correct answers, responses in more distant measures, and prolonged response times in any given measure. The data for the difference in the 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 stimulus pairs but two (MPR5a, d: *length, articulation*), the ratio of correct and incorrect responses significantly differs in expected and unexpected stimulus pairs. Stressing the same point from a different angle, Table 4 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 location of 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 change in sequences with unexpected suggestions.

In short, the reduced number of hits (with $p< .05$ statistical significance except for length and articulation), and prolonged average latencies (in all pairs but length and slur, 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 that factors inducing them to press the button in a particular measure were partly extramusical.

Table 4 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		

There is one more result suggesting how important expectancies are: the dramatic drop of the musicians' accuracy in the unexpected stimulus 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 (see section 3.3). Thus, trained musical professionals too seem to have constructed metrical Gestalten based on preference rules different from the ones targeted in the present study. While these have technically been "incorrect" responses based on our methodology, such choices need not at all be a consequence of their "lack" of musical understanding. Rather, their preferences may just have contradicted the author's (and partly also MPR5's) predictions. This was especially prominent in the "length" example, where many musicians seemed to prefer the shorter tone as the location of the stressed beat. This may be so because most real melodies actually start with an upbeat, where listeners do not really have difficulty in identifying the metrical structure, if only after a few seconds.

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)

Expectancies are thus a question that has to be considered in any investigation of metrical perception. In the present study, their influence, especially on musicians, was obvious. Yet, the preliminary conclusion appears valid stating that, with small variations, in our population metrical constraint ranking was a relatively stable phenomenon (not strongly correlated with either musical education or disrupted expectancies). Due to the several minor inconsistencies in this result between musicians and non-musicians, this result should be fine-tested in further studies.

4. Discussion

The metrical preference rules from the group MPR5 proposed in GTTM seem to have empirical validity. The constraints indeed appeared to differ in intensity, according to the test hypothesis. The exact ranking, however, remains unresolved, as the present sample size and stimulus design failed to account for the position of adjacent constraints in the entire population. It was still possible to make statistically valid generalizations for three *groups* of constraints (Appendix A).

What could be the reason behind such a ranking? As for the first category, 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 (louder) 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 was not surprising, either: although a higher-order musical factor, chord sequencing seems to be so important to western ears that both musicians and nonmusicians considered this suggestion very relevant for determining meter, especially if it was well-formed, as was the case in our example (plagal cadence I-IV-I). In the present design, this was the only stimulus pair that explicitly confronted two constraints (length and harmony). Even if it is true that these are “different order” preference rules, it turned out that harmony was the definite winner. In Gestalt psychology terms, confronted here was “proximity” with “figure/ ground”, where the latter seems to be clearly structurally more important in metrical perception, a result that might be given some consideration in further research.

The second statistically delineated group by strength consisted of three individual constraints: slur, pitch, and length. A stronger note and prominent harmony, that appeared in the first group, are partly differentiated from the melodic line and provide a strong impetus to the parser to segment the musical structure at that exact location. With slur and pitch change, however, there is no such “additional” factor. The parser rather concentrates on the melodic progression and must infer the meter during this process. The 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 yet fewer elements for the participants to count on while deciding on the stressed beat, which may have reflected on the constraint ranking. Subsequent discussion with some musicians revealed 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, the 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. The author of the study is partly to be held responsible for this result, as the musical example offered was indeed a bit more difficult, albeit almost exactly copied from GTTM p. 82, ex. 4.29 (the succession of two sixteen-note quadruplets and eight-note triplets in a 4/4 meter signature, at 100bpm, MPR5d, Figure 1). Yet, although the complexity of the stimulus and a slightly faster tempo may have played a role, the stimulus may have been 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 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. The hope remains that further research will fine-tune this result. In particular, such further work should look for additional ways to make the examples more similar to one another according to *other* (non-targeted) parameters as well, so that any results in the final hierarchy could not be potentially attributed to between-stimulus differences. In terms of the present study, the main such concern seems to relate to the “length” and “articulation” examples, where the notes may have been too long in the first case and too fast in the second. This remains a proposal for corrections in further work.

The two auxiliary hypotheses have been partly corroborated. In terms of hypothesis 1a, musicians did have better results than nonmusicians in eight examples out of twelve ($p < .05$), of which they scored better in all six stimuli from the ‘expected suggestion’ group. Yet in four stimuli, all from 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 not fulfilled. Whether this had to do with their lack of concentration while performing the task, with the strong general influence of expectancies as a limiting factor, or with the possibility that they actually responded to some “hidden” preference rules that our study did not target, remains to be further investigated.

Classified as the three macro-groups proposed above, the constraint rankings of musicians and nonmusicians were similar. Some caution is warranted here. If one attempts 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 that the group of musicians had fewer participants (30 : 90). The remaining five constraints are equally classified in the two groups. While this minor difference remains, the result still seems important: the identical constraint ranking with expected and unexpected suggestions in the entire population, and the *almost* identical ranking achieved by musicians and nonmusicians may together support the central hypothesis of the present research, suggesting the stability of GTTM metrical constraint hierarchies. If this tendency should be proven crossculturally, too, then the ultimate universalist aim of GTTM (which it shares with the generative enterprise in linguistics) may not remain so far-fetched: it might even turn out that metrical segmentation, rather than the oft-studied grouping, is the principal domain in which GTTM all but achieved its goal – to target deep, structural musical universals.

Finally, the tenets of the second auxiliary hypothesis seem to be true. In the entire population, the answer distributions differed in four stimulus pairs out of six, where the expected group had significantly more correct answers, and the response latency was significantly longer in four unexpected examples out of six. Both tendencies suggest a strong influence of expectancies on metrical perception, yet without major changes in the ranking of the constraints (except for the position of MPR5a, length, Appendix C). In the present 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 further studied in the future.

5. CONCLUSION

This study has attempted to show that the metrical preference rules from Group 5 proposed in GTTM can be experimentally corroborated. The results suggest that constraints may be ranked into three macro groups, although their precise, individual ordering remains a task for future research. When doing metrical segmentation tasks, musicians and nonmusicians differ in several respects (the number of hits, average latencies...), but their internal constraint rankings, generalized into three macro-groups, are similar. Finally, in metrical perception, expectancy remains an important construct, but it does not significantly influence the 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, this study has not *deliberately* confronted constraints in the same examples (except for MPR5f). Further research could also more deeply consider latencies and the exact measure in which the constraint was responded to as variables influencing the final ordering of the constraints. We approached these data in relation to the problem of expectancy alone. Likewise, in terms of constraint ranking, 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 the stimuli or a larger sample (or both). In calculating this ranking, though, the only factor that we took into account was the “correctness” of the response. Thus, in the present study, conflicts of constraints in individual examples were not explicitly studied. Another interesting proposal for future work would thus be to also consider the “*incorrect*” responses given by the participants and analyze whether they may not have been simple “mistakes” but rather results of the prevalence of preference rules *other* than the targeted ones. Finally, the big question of music cognition research remains in the end: the experimenter is always under pressure to fully simplify the stimuli for methodological reasons; at the same time, the simpler these stimuli are, the less ‘musical’ they sound and the question remains of whether the phenomenon studied has anything to do with realistic music at all. In the present study, the explicit focus on one small rule from GTTM has hopefully justified the use of relatively bare-bone metrical stimuli, but in future work more “musical” material, perhaps also coming from the actual classical repertoire, would be welcome. This would likely result in the simultaneous activity of a number of constraints in individual examples, and would in turn require a much more complex design, with many more stimuli and a full-scale factor analysis. As such, it remains an interesting proposal for future work.

The hope remains, however, that the present study, too, has shown that preference rules/constraints should be favored over binary choices, at least in the segmentation of metrical patterns by western ears. This itself was a remarkable prediction of the often praised, but also criticized, quarter of a century old theory of music cognition. In terms of metrical segmentation, GTTM seems to have stood the test of time.

REFERENCES

1. Clarke, E., & Krumhansl, C., (1990), "Perceiving musical time", *Music Perception, 7,* 213-251.
2. Deliege, I., (1987), "Grouping conditions in listening to music: an approach to Lerdahl and Jackendoff's grouping preference rules", *Music Perception, 4*, 325-360.
3. 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.
4. Gilbers, D., & Schreuder, M., (2002), "Language and music in optimality theory", *Rutgers Optimality Archive*, ROA 571-0103, http://roa.rutdgers.edu.
5. Grice, P., (1975), "Logic and conversation", In P. Cole & J. Morgan (Eds.), *Syntax and Semantics*, vol. 3, (pp. 41-58), New York: Academic
6. Hasty, C., (1997), *Meter as Rhythm*, Oxford University Press
7. Hamanaka, M., Hirata, K., & Tojo, S. (2013). "Computational Music Theory and Its Applications to Expressive Performance and Composition". In A. Kirke & E. Miranda (Eds.), *Guide to Computing for Expressive Music Performance* (pp. 205-234). Springer London
8. Huron, D. (2006), *Sweet Anticipation: Music and the Psychology of Expectation*, MIT Press.
9. Jackendoff, R., (2002), *Foundations of Language*, Oxford University Press
10. Jackendoff, R. (2009). "Parallels and Non-parallels between Language and Music", *Music Perception, 26, 3*, 195-204.
11. Jongsma, M., Quiroga, R., & van Rijn, C., (2004), "Rhythmic training decreases latency jitter of omission evoked potentials in humans", *Neuroscience Letters 355*, 189-192.
12. Large, E., & Palmer, C., (2002), "Perceiving temporal regularity in music", *Cognitive Science Journal*, *26, 11*, 1-37.
13. Lerdahl, F., (2001), *Tonal Pitch Space.* Oxford University Press
14. Lerdahl, F., & Jackendoff, R., (1983), *A Generative Theory of Tonal Music.* MIT Press
15. London, J. (2012), *Hearing in time: Psychological Aspects of Musical Meter*. Oxford University Press
16. Parncutt, R., (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception 11*, 409-464.
17. Prince A., & Smolensky, P., (1993), *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University Center for Cognitive Science Technical Report 2
18. Quene, H., & Port, R., (2005), "Effects of timing regularity and metrical expectancy on spoken-word perception", *Phonetica, 62*, 1-13.
19. Roberts, S., (1996). *Interpreting Rhythmic Structures Using Artificial Neural Networks.* Unpublished dissertation, University of Wales, College of Cardiff
20. Rohrmeier, M., & Koelsch, S. (2012), "Predictive Information Processing in Music Cognition. A Critical Review", *International Journal of Psychophysiology*, *83, 2*, 164-175.
21. Ross, J.R., (1970), *Constraints on Variables in Syntax,* Unpublished dissertation, Massachusetts Institute of Technology
22. Rothstein, W., (1989), *Phrase Rhythm in Tonal Music,* New York: Schirmer Books
23. Temperley, D., (2000), "Meter and grouping in African music: a view from music theory", *Ethnomusicology*, *44, 1*, 65-96.
24. Temperley, D., (2004), *The Cognition of Basic Musical Structures.* MIT Press
25. 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
26. 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)

LINGVISTIČKI KONSTRUKT POMAŽE MUZIKOLOGIJI: HIJERARHIJA METRIČKIH OGRANIČENJA PRI PERCEPCIJI MUZIKE

Konstrukt perceptivnih ograničenja poslednjih godina postaje sve značajniji u kognitivnim naukama, uključujući i istraživanja na granici lingvistike i muzikologije. Ovaj rad prikazuje rezultate empirijskog istraživanja rasporeda metričkih pravila izbora / ograničenja iz grupe MPR5, predloženih u "Generativnoj teoriji tonalne muzike" (GTTM, Lerdahl & Jackendoff, 1983). Teorija predviđa preferirano zaključivanje jakog udara na muzičkim tonovima koji proizvode relativno izraženu promenu tonske visine, dinamike, realtivno dugi niz vezanih nota, dugi artikulacioni sklop, dugo trajanje visine tona na relevantnom nivou redukcije vremenskih odseka, kao i istaknutu harmoniju na relevantnom nivou redukcije vremenskih odseka. Sto dvadeset slučajno odabranih studenata osnovnih studija (30 muzičara i 90 nemuzičara) slušalo je dvanaest metričkih sekvenci baziranih na primerima za pravilo MPR5 iz GTTM. Od toga, polovina primera bila je usklađena sa kognitivnim očekivanjima ispitanika, a polovina se protivila istim očekivanjima. Učesnici su dobili instrukciju da pritisnu dugme kada budu sigurni da su čuli naglašeni udar. Distribucije odgovora sugerišu da šest ograničenja može da se grupiše u tri veće kategorije, i to: (*dinamika, harmonija*), (*tonska visina, ligatura, trajanje*) i (*artikulacija*). *Muzičari su postigli bolje rezultate nego nemuzičari, a vreme reakcije je značajno poraslo u stimulusima koji su bili suprotstavljeni očekivanjima, ali interne hijerarhije ograničenja ostale su relativno stabilne bez obzira na ta dva faktora (muzičko obrazovanje i početak stimulusa na ciljanom udaru). Imajući u vidu takve rezultate, postavlja se hipoteza da je metrička segmentacija glavni doprinos GTTM koji je izdržao proveru vremena.*

Ključne reči: *muzički metar, pravila izbora, ograničenja, teorija optimalnosti, generativna teorija tonalne muzike.*