Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009) Jyväskylä, Finland
Jukka Louhivuori, Tuomas Eerola, Suvi Saarikallio, Tommi Himberg, Päivi-Sisko Eerola (Editors)

Grammar Types in Language Explain Tone Sequence Processing in Music

Christiane Neuhaus, Thomas R. Knösche, Jörg Bahlmann, Angela D. Friederici

Max Planck Institute for Human Cognitive and Brain Sciences, Germany

`neuhaus@cbs.mpg.de, knoesche@cbs.mpg.de, bahlmann@cbs.mpg.de, angelafr@cbs.mpg.de`

ABSTRACT

In this ERP study, linear and center-embedded musical sequences are built according to two artificial grammar types in language, named finite state grammar (FSG) and phrase structure grammar (PSG). The aim is to prove if neural sources and processing mechanisms for artificial grammar settings across domains are the same. Isochronous pitch sequences were constructed by two interval categories (3rd and 6th) in upward and downward direction. FSG sequences, which have the general form ABAB in artificial grammar, are translated into "small up/small down/large up/large down". PSG sequences of form A[AB]B are transposed to "small up/large up/large down/small down". In two ERP recordings testing FSG and PSG separately, non-musicians had to distinguish between correct and false examples after getting familiar with each grammar type. Deviant sequences either include an item of reverse interval or contour. Our main results are: (1) N1 components indicate a 2-item-chunking in FSG and a 4-item-chunking in PSG based on immediate repetition between adjacent tones, thus low-level grouping is different for each grammar type. (2) A late processing negativity at sequence offset indicates syntax-based integration-and-memory processes primarily for PSG. The partially congruent ERP results for artificial grammar learning in language and music confirm that the linguistic perspective on music may be justified.

I. INTRODUCTION

Since the seminal work of Noam Chomsky, researchers draw parallels between music and language and seek common cognitive principles in both domains. Over the past two decades there has been an ongoing tendency to explain syntax processing in music by concepts in linguistics. Strong evidence has been found that processing mechanisms in both domains are partly the same and that, functio-anatomically, the same brain regions are activated. For the ease of handling this topic, let us concentrate on the aspect of syntax and leave aside the interactions with phonological and semantic effects.

Three relevant insights into the processing of musical and language syntax should briefly be mentioned here. First, Maess and colleagues (2001) proved by MEG source localization that syntactic incongruities in music, in particular unexpected chords in musical cadences, are processed in Broca's area (BA 44/45) which was previously considered as specific to syntax processing in language (see also Koelsch et al., 2002, for comparable results using fMRI). Second, Levitin and Menon (2003) showed in an fMRI study that tone sequences of temporal coherence as compared to disrupted and random musical structures are mainly processed in the pars orbitalis region (BA 47) of the left inferior frontal cortex which was previously identified as the circumscribed brain region for verb generation and semantic word processing. Third, Patel (1998, 2003) suggests a shared cognitive mechanism for processing musical syntax and language syntax, called "shared structural integration resource" hypothesis. According to this hypothesis, two locally separate components are located in anterior and posterior parts of the brain, the former is assigned to allocating resources for integrating and memorizing lexical items, the latter is assigned to syntax representation and storage.

Let us focus on center-embedded (nested) structures in language as compared to music and distinguish between natural and artificial types of grammar. In natural languages, such as English or Italian, center-embedded structures bring about hierarchical relations between subordinate items of the relative clause and the respective anchor words in the main clause which leads to distance-based dependencies in word structure (Gibson, 1998). In music, a substantial portion of structure reveals similar hierarchical relations, most evident in chord progressions with the tonic center as the main reference (cf. Patel, 2003).

As several attributes in natural languages, for example animacy and case, play a substantial role in processing grammatical relations, it is, of course, a lot easier to explore the processing of center-embedded structures in artificial grammar learning tasks in which the only modified variables are category and word order. In artificial grammar learning, Friederici and Bahlmann (2006) studied long and short consonant-vowel sequences (e.g. "le ri se de ku bo fo mo") based on two grammar types called finite state grammar (FSG) and phrase structure grammar (PSG). FSG is determined by local transition probabilities. Sequences follow the rule $(AB)^n$, and the resulting structure is ABAB (e.g. "de bo gi fo"). PSG is characterized by recursive embeddings and accordingly, long-distant dependencies are processed. PSG sequences follow the rule (A^nB^n), yielding the structure AABB (e.g. "de gi bo fo"). Thus, the study compares linear with this hierarchical sequence structure, and A and B categories are distinguishable from each other by bright and dark vowels.

The present study uses the artificial grammar learning paradigm by Friederici et al. (2006) and translates it to music at a ratio of 1:1. Our objective is to prove if FSG and PSG grammar types in language can describe (pitch) structure in music appropriately, and if neural sources and processing mechanisms are once again similar across domains so evidence towards domain-general processing is accumulating. For modelling this kind of grammar setting in music, we use isochronous pitch patterns, each consisting of eight tones in two interval categories (3rd and 6th) and two pitch directions (up and down). The finite state grammar (FSG), as illustrated by structure ABAB, is translated into the pitch sequence "3rd up/3rd down/6th up/6th down" whereas condition PSG, as represented by structure AABB, is transposed into "3rd up/6th up/6th down/3rd down" (see note examples in Figure 2). The underlying principle in both sequences is therefore an alternation of openings and closures. FSG and PSG tone sequences were presented in two separate sessions, and event-related potentials (ERPs) were recorded while listening to sequence examples. Subjects had to distinguish between correct (standard) and false (deviant) stimulus examples in each grammar type.

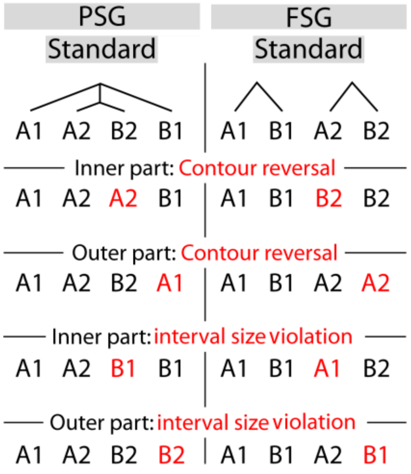

Figure 1. Schematic diagram illustrating PSG and FSG sequence types (standard and deviant forms).

II. METHODS

A.Subjects

15 students from the University of Leipzig participated in the experiment. Age, gender, and their field of study were balanced (7 males, 8 females; average age = 24.3 years, *SD* = 2.4). According to a filled questionnaire asking for the type of instrument, playing practice, and score-reading abilities, all participants were specified as non-musicians. Each student was tested in two alternate sessions, one for FSG, the other for PSG.

B.Stimuli and Task

FSG and PSG pitch sequences consist of eight isochronous tones, each with a length of 250 ms, no pause in-between. Thus, overall sequence length is 2 s in all examples. Each sequence is built on two interval categories (3rd and 6th) that appear in two pitch directions (up and down). Sequences that follow the finite state grammar rule (FSG) are of the basic structure "interval1 up/interval1 down/interval2 up/interval2 down", whereas sequences built according to the principles of phrase structure (PSG) have the basic form "interval1 up/interval2 up/interval2 down/interval1 down". To avoid pattern learning as much as possible, i.e. to let participants focus on the rule-based aspect, we systematically varied major and minor types of each interval category when building sequences. Accordingly, FSG standard patterns appear in eight different forms, and PSG standard patterns are built in a similar manner. To perform the task ("identify the correct learned sequence"), we contrasted FSG and PSG standards with four deviant sequence types which either included an item of opposite interval size (e.g. major 3rd instead of major 6th) or of reverse contour in inner or outer parts of the second pattern half (cf. note examples in Figure 2).

Pitch sequences were generated as MIDI files in the timbre 'acoustic piano' and transformed to audio format (wav) for automatic presentation (Experimental Run Time System, Version 3.11, BeriSoft 1995). Stimuli of the same grammar type, FSG or PSG, were presented in five blocks, each consisting of 64 stimuli. Within each block, sequence order was pseudo-random, and the number of standard forms was four times higher as compared to each deviant alone. Subjects indicated their rating decisions by press of key buttons immediately after stimulus presentation. They were also requested to keep head, neck, arms, hands, and fingers as relaxed as possible and to reduce the amount of eye blinks during recording. Each recording session started with a learning phase up to three test blocks of ten standard trials to learn the respective grammar type and get familiar with the task. During the experimental run, feedback was given after each block.

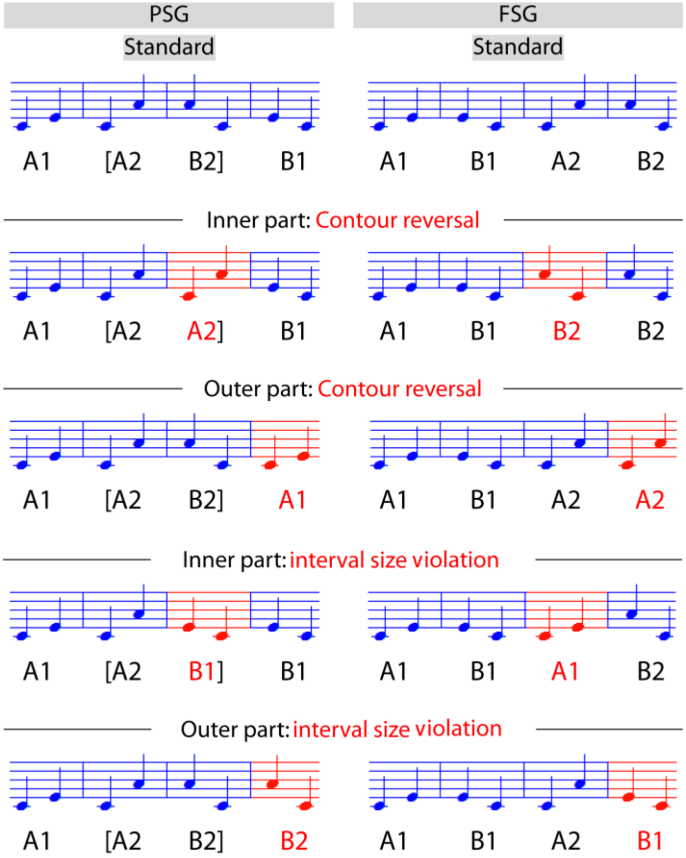

Figure 2. Pitch sequences following artificial grammar rules. Left: Center-embedded structures according to PSG. Right: Linear structures according to FSG. Top row: Standard types. Middle and bottom: Deviant sequences including reverse contour or interval items in the inner and outer parts of the second pattern half.

C.Recording Procedure and Preprocessing of Data

We used standard procedures to measure brain electrical activity and for preprocessing EEG raw data so description will be kept short. 59 active Ag/AgCl electrodes were placed onto the head's surface according to a modified version of Jasper's 10-20 system (Oostenveld & Praamstra, 2001). Channel activity was recorded with an infinite time constant and referenced to the left preauricular point (A1). The sternum was used for the ground electrode. Ocular artefacts were measured with a vertical and a horizontal electrooculogram (EOGV, EOGH). Data sets with EEG raw signals were first high-pass filtered (cut-off frequency of 0.50 Hz) and examined for eye blinks and muscle artefacts using an automatic algorithm. Artefact-free trials were then merged over blocks and examples, but averaged separately according to grammar type (FSG vs. PSG), condition (standard, four deviants in terms of interval or contour), and channel. Averaging was done for a time window between -200 and 2500 ms measured from sequence onset, and resulting ERP traces were baseline-corrected (time interval: -50 to 0 ms). Curves were finally subsumed to overall grand average ERPs (cf. Figures 3 and 4).

D.Statistical Analysis

To validate the results of the visual data inspection, we computed three separate types of ANOVAs (repeated-measures analyses of variance). Each analysis refers to one component (N1, P3, and late processing negativity) in the respective time window. Repeated measures factors were Grammar Type (FSG, PSG), Condition (standard, four deviants), Window (P3 interval, P3 contour), and Channel (36 electrodes). Degrees of freedom were adjusted with Huynh and Feldt's epsilon, and results were considered significant at an alpha level of .05. Detailed results are listed in Table 1.

III. RESULTS

To explain the results of this study, we refer to the ERP traces in Figures 3 and 4. Figure 3 shows the grand average ERP for the FSG standard type as compared to the PSG standard. Figure 4 illustrates how center-embedded deviant structures (PSG deviants) are processed as compared to the PSG standard form. In Figure 3, we observe a clear N1 component on every first tone of each interval category in the FSG standard curve, i.e. for brain responses to tones 1, (2), 3, 5, and 7. The brain pattern for processing PSG standard sequences looks different. Here we observe a distinct N1 on every fourth tone, i.e. for brain responses to the 1st, (2nd), and 5th tone. To confirm results for the different N1 amplitudes between FSG and PSG conditions, we computed a two-factor ANOVA (Grammar type, Channel) for each sequence tone separately, thus, getting eight different results. Statistical analysis yields a significant main effect of Grammar type (FSG vs. PSG standards) for tones 3, 4, 6, and 8 (cf. Table 1), whereas for the remaining tones (1, 2, 5, and 7) results are non-significant, so amplitude similarity between FSG and PSG is validated in these cases.

Let us now discuss how deviant PSG sequences, i.e. violations of center-embedded structures, are processed as compared to the PSG standard form. Figure 4 reveals the two main results, a (centro-)parietal positive shift (P3) for inner-part interval and contour deviants, and a clear fronto-centrally distributed negativity immediately at sequence offset (from 2 s onward) for three deviant conditions and the standard form.

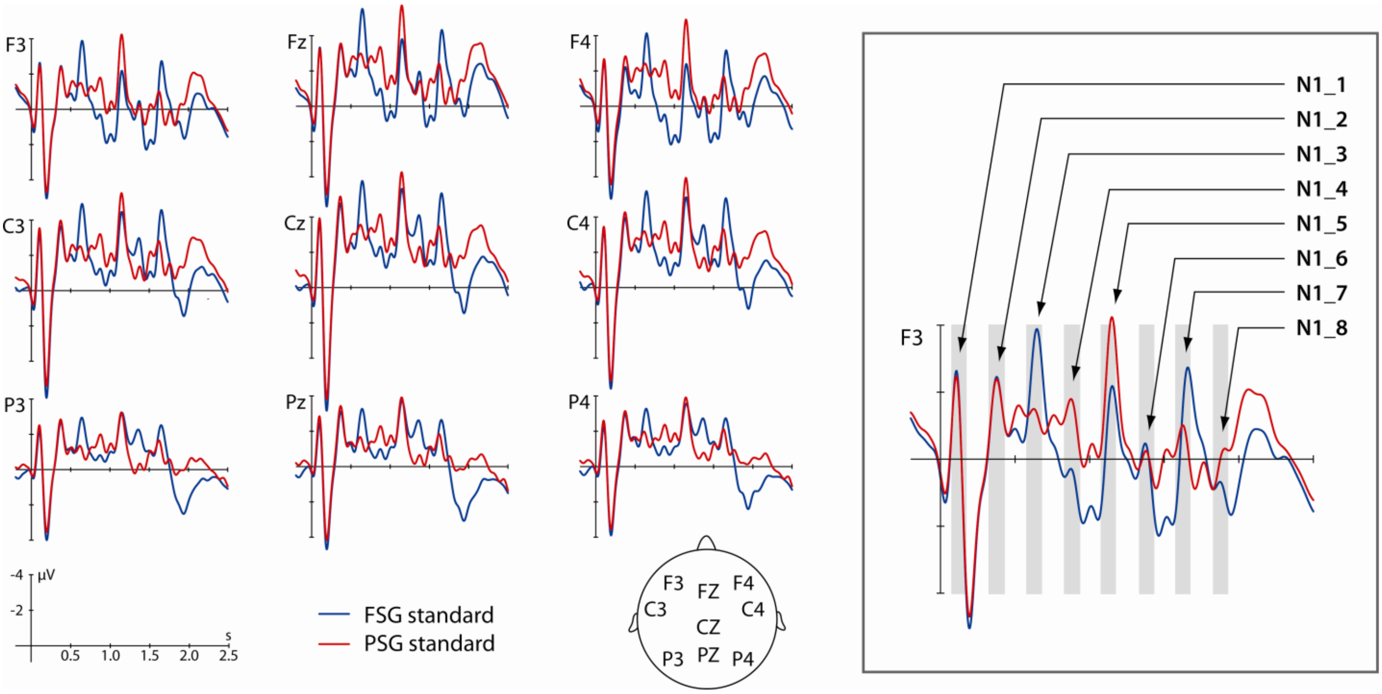

Figure 3. Grand average ERPs, nine selected electrodes. Brain response to entire sequences (sequence length 2.0 s).

In terms of the P3, closer inspection reveals a latency shift at parietal electrode sites, in that reverse intervals are processed earlier than reverse contour. For this early versus delayed processing of deviant forms in the inner part of PSG sequences, ANOVA yields a significant main effect of Window in the respective time ranges (1.46 to 1.77 s for reverse interval, and 1.64 to 1.96 s for reverse contour, cf. Table 1). Let us finally discuss the main characteristics of the late anterior negativity at sequence offset, in the following called late processing negativity. The component reveals a clear scaling in amplitude size which is largest for reverse contour in the inner part of the second half of the sequence, closely followed by smaller negativities for inner-part interval size violation and for the standard form, and by an even smaller negativity for processing outer-part interval violation (cf. Figure 4).

However, the most obvious brain response is elicited by reverse contour in the outer part of the sequence (cf. Figure 4 and note examples in Figure 2). Here we assume that, due to the change in pitch direction, it appears perceptually as if after three interval items the sequence starts again. This assumption is supported by the fact that strong N1 and P2 components can be observed from 1.6 s onward, usually indicating a regular onset.

In terms of the late processing negativity, ANOVA yields two significant results in the time range from 2.0 to 2.35 s. First, a significant main effect of Grammar Type, providing evidence for the late processing negativity existing in PSG as opposed to FSG. Second, a significant main effect of Condition which confirms the scaled characteristic of this component.

FSG deviants as compared to the standard form do not elicit any comparable components like those shown in Figure 4, neither a parietal P3 nor a scaling of the late processing negativity which in FSG is much smaller than in PSG (not displayed because of space limitations).

IV. DISCUSSION

So far, pitch sequences built according to artificial grammar rules are processed in three distinguishing ways, indicated by the N1, P3, and the late processing negativity. We now try to explain the assumed underlying cognitive principles.

Visual inspection of the ERPs reveals a different distribution of N1 components for FSG as compared to PSG sequences, i.e. for the 1st, (2nd), 3rd, 5th, and 7th tone in FSG and for the 1st, (2nd), and 5th tone in PSG (cf. Figure 3). In both conditions, peak amplitudes may not originate from processing pitch height as a purely acoustic feature alone as tones evoking this component are either the tonic or the third and the sixth note of the respective interval (see note examples in Figure 2). ERP traces thus suggest a different perceptual grouping mechanism for each grammar type, resulting in 2-item chunks for FSG and 4-item chunks for PSG. However, although we can exclude that encoding takes place on the basis of mere acoustic features, grouping may simply be caused by identifying tone repetition between adjacent sequence tones which results from the chosen paradigm (repetition is at position 2./3., 4./5., and 6./7. in FSG and at position 4./5. in PSG). When we compare note examples and ERP traces from this proximity-based point of view (cf. Figures 2 and 3), this explanation indeed seems

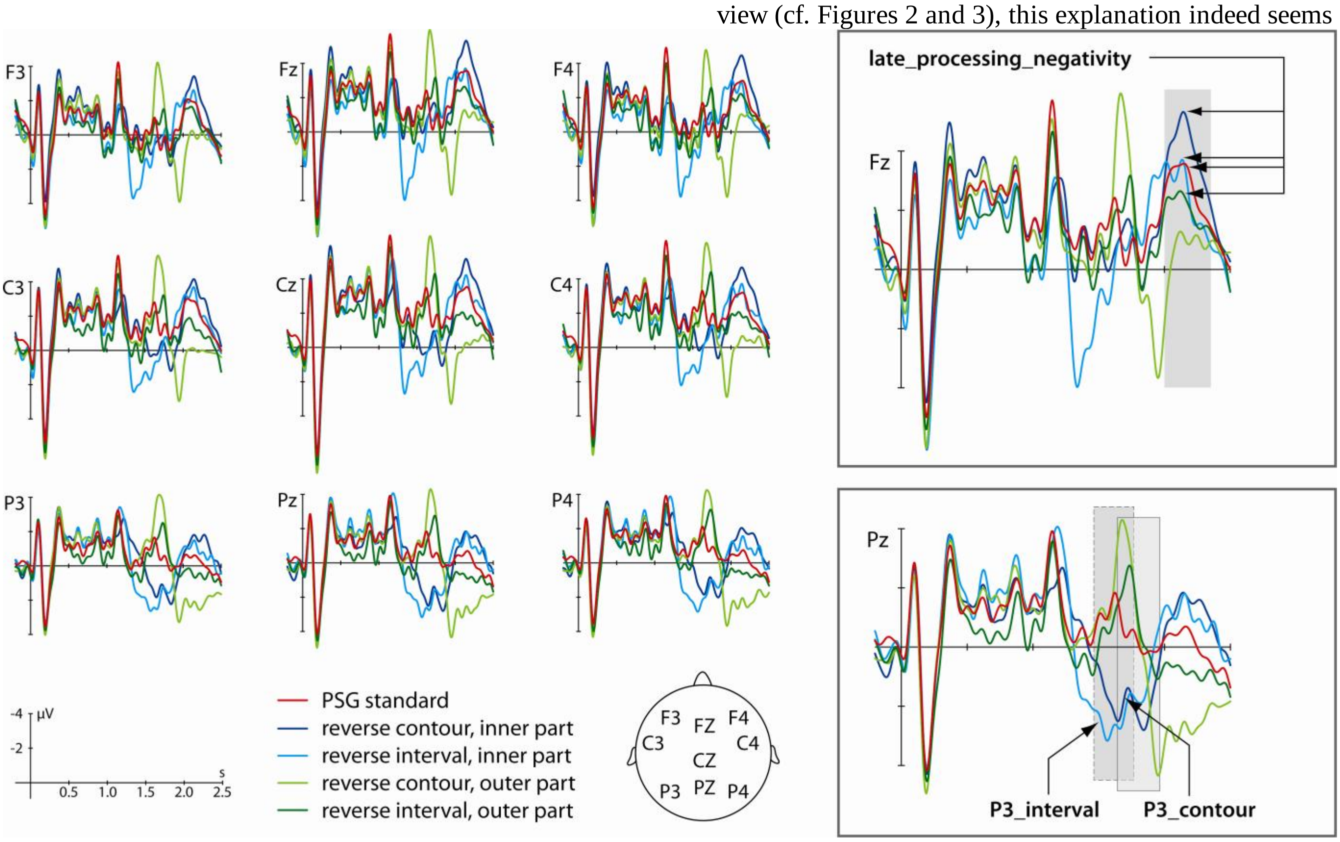

Figure 4. Grand average ERPs for entire PSG sequences (sequence length 2.0 s, standard and deviant forms).

to be plausible. So with regard to N1 results we can abandon any assumption of top-down, syntax-based grouping, as N1 results can be fully explained by a repetition-based low-level grouping process.

The second result of the study refers to the P3 at parietal electrode sites. It is obvious that whenever violations of interval size or contour occur in the inner part of the second sequence half, the brain's reaction is a parietal P3. The significant latency shift between these deviant conditions indicates that local processing of interval size is prior to contour. This result is roughly in line with previous research, although temporal order of local interval and global contour effects seems to be different in preceding studies as compared to the present one. Trainor et al. (1999), for example, could demonstrate that the parietal P3b remains constant across subjects (musicians vs. nonmusicians) whenever contour violations (rising contours) are processed, but is delayed and smaller in nonmusicians whenever interval violations are perceived, which is a slight contradiction to our results.

Lastly, let us turn to a possible interpretation for the late processing negativity revealing a scaled structure. For the PSG as compared to the FSG condition, amplitudes are higher which, along with the extended distribution of brain activity from frontal to parietal areas, suggests higher processing demands for PSG sequences. So a syntax-based origin of the late processing negativity seems likely, and the component probably refers to the entire sequence rather than to processing the final item within the sequence context.

Furthermore, our results for center-embedded vs. linear sequence structure (FSG vs. PSG) are in general accordance with the dual-syntax approach elaborated by Patel (2003), suggesting syntax processing and syntax storage to be located in different, i.e. anterior and posterior regions of the brain. According to these ideas, the (anterior) late processing negativity in the present study, with the scaled form as the distinctive feature, might indicate higher processing costs for center-embedded as compared to linear sequences, in particular for the processing of inner-part violations. However, the present study does not allow a clear differentiation between integration and memory aspects, i.e. the two sub-components of syntax processing as proposed by Patel (1998, 2003).

When drawing parallels between artificial grammar processing in language (Bahlmann et al., 2006) and music, there is a certain overlap between results, especially in terms of the parietal P3 which was interpreted as reflecting some difficulty to integrate deviant categories into the syllable context (e.g. "ti me **pe** gu" instead of "ti me fo gu") However, the language study conducted by Bahlmann et al. (2006) differs from our music study in such a way that neither distinct N1s nor a late processing negativity could be identified. The missing N1 in the language domain may be caused by not presenting syllable sequences in an acoustical manner but rather in a visual slide-by slide mode. A missing late processing negativity may indicate less effort to process center-embedded (nested) structures in language as compared to music, thus hierarchical syntax constructions seem to be more natural in language than music, especially when new input has to be matched against stored templates in long-term memory.

Table 1. Repeated measures ANOVAs for artificial grammar processing in music.

N1 Amplitude		
FSG Standard - PSG Standard		
Two-Factor, per Time Window (Grammar Type, Channel)		
Source	**df**	**F value**
Grammar Type (ms)	**1,13**	
80 - 180 (tone 1)		0.4 (n.s.)
330 - 430 (tone 2)		0.03 (n.s.)
580 - 680 (tone 3)		14.6 **
830 - 930 (tone 4)		8.44 *
1080 - 1180 (tone 5)		2.58 (n.s.)
1330 - 1430 (tone 6)		9.02 **
1580 - 1680 (tone 7)		3.05 (n.s.)
1830 - 1930 (tone 8)		4.6 (*)
P3 Latency		
PSG, testing 2 Deviant Forms only (Inner Part Interval vs. Inner Part Contour)		
Two-Factor (Window, Channel) (1.46 - 1.77 and 1.64 - 1.96 s)		
Source	**df**	**F value**
Window (P3 Interval vs. P3 Contour)	**1,14**	6.52 *
Late Processing Negativity		
1. FSG - PSG		
Two-Factor, per Time Window (Grammar Type, Channel)		
Source	**df**	**F value**
Grammar Type	**1,15**	10.45 **
2. Standard vs. Deviant Forms, only PSG condition		
Two-Factor, per Time Window (Condition, Channel)		
Source	**df**	**F value**
Condition	**4,56**	5.54 **

$^{(*)}$p < .06 *p < .05 **p < .01 ***p < .001

V. CONCLUSION

In summary, ERP results for artificial syntax processing partly overlap between music and language which again confirms the linguistic perspective on music.

ACKNOWLEDGMENT

We thank Heike Böthel for careful EEG measurements.

REFERENCES

Bahlmann, J., Gunter, T. C., & Friederici, A. D. (2006). Hierarchical and linear sequence processing: an electrophysiological exploration of two different grammar types. *Journal of Cognitive Neuroscience 18(11)*, 1829-1842.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *PNAS 103(7)*, 2458-2463.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition 68*, 1-76.

Koelsch, S., Gunter, T. C., v. Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical „language-network“ serves the processing of music. *NeuroImage 17*, 956 - 966.

Levitin, D. J., & Menon, V. (2003). Musical structure is processed in “language” areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *NeuroImage 20*, 2142-2152.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca’s area: an MEG study. *Nature Neuroscience 4(5)*, 540-545.

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology 112*, 713-719.

Patel, A. D. (1998). Syntactic processing in language and music: different cognitive operations, similar neural resources? *Music Perception 16(1)*, 27-42.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience 6(7)*, 674-681.

Trainor, L. J., Desjardins, R. N., & Rockel, C. (1999). A comparison of contour and interval processing in musicians and nonmusicians using event-related potentials. *Australian Journal of Psychology 51(3)*, 147-153.