Universality in structural grouping of simple melodies

Rebecca Schaefer, June-August 2003
Department of Psychonomics, perception
Psychology Department, University of Amsterdam

Supervision
Prof. Dr. J.M.J. Murre,
Psychonomics, Psychology Department, University of Amsterdam
Dr. R. Bod,
Institute for Logic, Language and Computation, University of Amsterdam

Abstract
We review parsing of simple melodies in music perception. Bod (2001) ascertains a phenomenon in a coded musical database that cannot be explained by either Gestalt Principles of perception or musical factors, suggesting a memory-based approach to musical analysis, in which new pieces of music are analysed by combining fragments from structures of previously encountered pieces. In the current experiments we will try to replicate the musical segmentation that gives cause to theorise on exceptions in the Gestalt principles of perception. We will review whether the way of segmenting of simple melodies is universal for different levels of musical experience, and if the same phenomenon will be seen as is present in the available coded database. In the current experiment, 30 subjects of varying degrees of musical experience segmented simple folk melodies. The data were reviewed, concentrating on the universality of the parsing process as well as the generalizability of the parses available in the musical dataset.
The results show that highly experienced listeners are more consistent in their parsing process, significantly differing from inexperienced listeners. This is most likely caused by an automatic, memory-based perception of melodic structure. Both the moderately and highly experienced group showed substantially more exceptions to the Gestalt principles in their manner of parsing. It appears that the Gestalt principles are not as generally applicable to melodic perception as it is in visual perception. The segmentation process for simple melodies appears to be mediated by the presence of musical experience, supporting the notion of a memory-based mechanism of music analysis.

With thanks to Neta Spiro for many helpful comments, and Sam Franklin for the technical support.

Introduction

The structural grouping mechanism in perception processes has a hierarchical nature. A clear example is language, where the stimulus is divided into sentences, and the sentences are structured according to the nature of the intrinsic information. In music perception, the listener segments a melody into groups of notes that form the structure of a usually large stimulus (e.g. Deutsch, 1980, Lerdahl & Jackendoff, 1983). In computational modelling of this parsing process ambiguities are found. These appear in melodies for which several ways of grouping are possible, whereas the listener typically prefers only one of these possibilities. The assumption is that this preference is based on low-level phenomena, such as local discontinuities and intervallic distances, and high-level phenomena like melodic parallelism and internal harmony. In testing a memory-based approach to musical analysis, in which new pieces of music are analysed by combining fragments from structures of previously encountered pieces, Bod's (2001) qualitative analysis of the Essen Folksong Collection, described below, indicates that there are grouping phenomena that challenge the commonly accepted Gestalt principles of perception. Such a phenomenon is seen in tonal phrases that include a jump (a large pitch interval) at the beginning or end of a phrase, when the interval to the closest note in the adjacent phrase is much smaller, from here on referred to as jump-phrases. This challenges the Gestalt principles of proximity and similarity, provided that this manner of grouping can't be explained by musical factors such as parallelism. Bod (2001), in proposing a memory-based system of musical analysis, makes the assumption that this analysis is based on the occurrence frequency of musical segments, and that excerpts (or parts thereof) that are familiar to the listener are grouped in the same way as in the previously encountered piece. This would result in a grouping process or mechanism that is influenced by the amount of experience one has with music.

It is widely assumed that musical training has an effect on musical perception due to a more thorough understanding of structure and categorisation (cf. Sloboda, 1985). In this light, the annotation of the Essen Folksong Collection may not represent a universal perception process. Bod (2001) states that the correctness of an annotation should preferably be established by an independent psychological experiment with more than one subject. The current research aims to ascertain the general universality of the melodic parsing process with musical training as an independent variable, focusing on the possible exceptions to the Gestalt principles.

Grouping in music theory

In their Generative Theory of Tonal Music (GTTM) Lerdahl & Jackendoff (1983) formulated a generative-linguistic 'grammar' for music. Combining the formal methodology and psychological concerns of Chomskian linguistics with the insights of Schenkerian music theory, the authors describe how a listener, experienced in the tonal idiom, intuitively creates an understanding of a complete musical structure. In this theory, the authors adopt two idealizations; that of the 'experienced listener' and that of 'final understanding'. The 'experienced listener' refers to the equivalent of Chomsky's 'ideal speaker-listener', and is introduced into the theory as to exclude listeners without sufficient experience to the western tonal idiom, assuming that the more experience the listener has, the better their theory will

explain their perception process. The concept of ‘final understanding’ refers to the difference between structure perception during real-time listening and structure perception after someone has become familiar with a musical piece.
Restricting themselves to those aspects of musical intuition that they consider hierarchical in nature, Lerdahl and Jackendoff (1983) divide their theory into four main components: grouping structure, metrical structure, time-span reduction, and prolongational reduction. Grouping structure is the most important component for this experiment; this is the part of the theory we will focus on.

Grouping structure is the most basic component of musical understanding, expressing a hierarchical segmentation of the piece into units such as motives, phrases, and sections. The intuitively preferred manner of grouping is stated in well-formedness rules, which specify the possible structural descriptions, and in preference rules, which state the listeners grouping preference, mostly based on the Gestalt principles of perception. Examples of grouping preference rules (GPR’s) are:

- GPR 1: Strongly avoid groups containing a single event.
- GPR 6: Where two or more segments of the music can be construed as parallel, they preferably form parallel parts of groups.

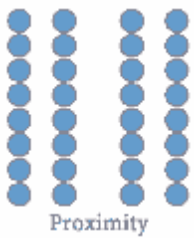
(Lerdahl & Jackendoff, 1983). For a schematic overview of the GTTM, see Appendix A.

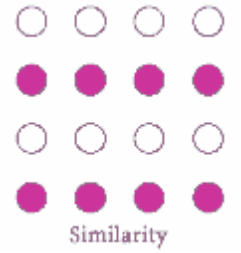
Grouping in Gestalt theory

Gestalt theory first arose in 1890 as a reaction to the prevalent psychological theory of the time, namely atomism. Atomism examined parts of things with the idea that these parts could then be put back together to construct a whole. Atomists believed the nature of things to be absolute and not dependent on context. Gestalt theorists, on the other hand, were intrigued by the way our mind perceives wholes (or ‘Gestalts’) out of incomplete elements. They believed that context was very important in perception, and they outlined several fundamental and universal principles (sometimes even called 'laws') of perceptual organization. Grouping of stimuli is a fundamental aspect of this organization.
Derived from the study of visual perception, Gestalt theory is also used in understanding music perception, because similar processes could govern perception of images and of sounds. Gestalt principles relevant to music include:

- Proximity: Items close to each other tend to be perceived as grouped together.
- Similarity: Items that are similar in some way tend to be perceived as grouped together.
- Common Direction: Items tend to be perceived as grouped if they have a similar direction or motion.
- Simplicity: Items tend to be perceived in the most regular, symmetrical, or simple manner.

For visual examples of the first two principles, see the figures next to this text. In the left-hand example, the principle of proximity is shown in how one tends to perceive two double columns of dots instead of four separate columns, caused by their proximity to one another. In the right-hand example, the similarity in colouring causes us to see four rows of dots instead of columns or separate dots. These tendencies seem to be universal in that no matter how much experience one has with visual images, this is usually the way they are perceived.





According to Gestalt theory, listeners perceive a melody as a Gestalt according to these same principles. Three characteristics of Western tonal melodies relate directly to Gestalt principles and serve as perceptual organizers for that style: propinquity (proximity of pitch interval), repetition in pitch and rhythm (similarity), and finality (good continuation or common direction). What the Gestalt principles of perceptual organization suggest is that we may be predisposed towards interpreting ambiguous stimuli in one way rather than another by universal principles. Note that this is also a key element in Lerdahl & Jackendoff's (1983) preference rules.

Universality

The above stated Gestalt principles are widely accepted and believed to apply to everyone without exception. As stated earlier, the amount of experience one has with visual images would not seem to cause someone to group visual stimuli differently. But is this also the case for music? Cognitive theory leads us to believe that all structural organization is subject to one's reference material or schemata. This implies that structural organization is influenced by experience. So it seems debatable whether the use of Gestalt principles in the GTTM is appropriate, and whether these principles are as generally applicable in music as they seem to be in visual perception.

The Essen Folksong Collection and ESAC coding

The Essen Folksong Collection contains a large sample of (mostly) European folksongs that have been collected and coded under the supervision of the late Dr. Helmut Schaffrath at the University of Essen (see Bod 2001, for references). Currently, 6251 folksongs are publicly available at http://www.esac-data.org, although the total number of folksongs in the collection is reported to be over 20.000. Each Folksong is annotated with the Essen Associative Code (ESAC) which includes pitch and duration information, meter signatures and explicit phrase markers. No text is available. The pitch encodings in the Essen Folksong Collection resemble 'solfege': scale degree numbers are used to replace the movable syllables 'do', 're', 'mi', etc. Thus 1 corresponds to 'do', 2 corresponds to 're', etc. Chromatic alterations are represented by adding either a '#' or a 'b' after the number. The plus (+) and minus (-) signs are added in front of the number if a note falls respectively above or below the principle octave (thus –1, 1 and +1 all refer to 'do', though on different octaves. Duration is represented by adding a period or an underscore to the number. A period ('.') increases duration by 50% and an

underscore ('_') increases duration by 100%. If a number has no duration indicator its duration corresponds to the smallest value (which is found in the KEY field preceding each folksong, along with the key it should be played in). A pause is represented by 0, possibly followed by duration indicators. No loudness or timbre indicators are used in the Essen Associative Code. The phrase boundaries are indicated with a hard return. This constitutes a lack of hierarchical structure, both phrase-internal structure (as subphrases and motives) and phrase-external structure (as periods and subsections) are neglected (Bod, 2001).
For this reason subjects in this experiment were asked to assign phrases as short as seemed reasonably possible to them, thus ignoring hierarchical structure. When comparing the assigned boundaries to the given phrase boundaries in the Essen Folksong Collection, this provides us with more data, whereas for this comparison a surplus of phrases does not influence the results. The implications of this request for the other aspects of data-analysis will be discussed later.

In the the above-mentioned jump-phrases (meaning phrases that include a jump (a large pitch interval) at the beginning or end of a phrase, when the interval to the closest note in the adjacent phrase is much smaller) several Gestalt principles are violated. In the following example (Figure 1) the beginning of song K0059 from the Essen Folksong Collection is shown. The Essen group placed boundary points between the last two notes of the second measure, as well as the last two notes of the fourth measure. These notes being identical in both pitch and duration, the Gestalt principles would dictate their being grouped together. In this case, the principles of proximity, similarity and even good continuation and parallelism seem to apply. However, the Essen group has most probably viewed these notes as pick-up notes, one to end one phrase and the other to start up the next. This judgement is most likely based on a certain idea of internal structure. In the ESAC, the melody in this example would be notated as:

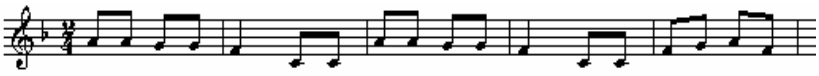


Figure 1: The opening measures of song K0059 from the Essen Folksong Collection

(33221_-5)(-533221_-5)(-51231…

with brackets used to delineate the assigned phrases. An alternative phrasing that would not violate the Gestalt principles would be

(33221_-5-5)(33221_-5-5)(1231… or (33221_)(-5-533221)(-5-51231…

The extent of ambiguity in phrase boundaries differs, and has to be assessed according to the different properties of the surrounding notes and patterns of notes. For example, Figure 2 shows another beginning of one of the folksongs, K0121B, where the first boundary point would most probably be placed at the end of the second measure. Not only is there a pause in the music, but also a substantial jump in pitch. So this can be considered a non- ambiguous phrase

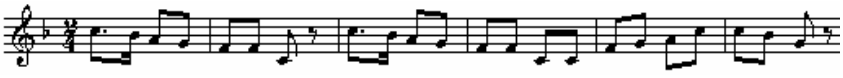


Figure 2: The opening measures of song K0121B from the Essen Folksong Collection

boundary. The Essen group placed the second boundary point between the last two notes of the fourth measure, which at first sight also violates the Gestalt principles of proximity and similarity, but on a closer look is explained by melodic parallelism. Because the previous phrase ended with only one low note, the next phrase is likely to be parallel, and be assigned as such. This leaves the last note of the fourth measure to be grouped in the next phrase. The above illustrates that the amount of ambiguity has to be assessed for a combination of notes within their surroundings, treating the total melody as a Gestalt.

Methodology

In the experiment as described in the following, several methodological choices have been made. To avoid an effect of complexity in the melodies, most of the test stimuli have been selected from the children's songs in the Essen Folksong Collection. However, to see the extent of this effect and to increase generalizability, a few tunes were also used that were more complex. This should not pose a problem when it comes to assessing universality in the parsing process, but it will be kept in mind when analysing the data.

Also, all stimuli selected contained one or more jump-phrases, which include a jump (a large pitch interval) at the beginning or end of a phrase, when the interval to the closest note in the adjacent phrase is much smaller. The extent of ambiguity in these phrases is not the same for all jump-phrases. Not only pitch interval, but other factors, like rhythm and meter (low-level phenomena) and melodic parallelism (high-level phenomenon) influence this extent of ambiguity. For this reason, jump-phrases and other phrase boundaries with varying levels of ambiguity were selected as test stimuli.

Finally, to assess a possible memory-based effect of the other melodies on the parsing process, the stimuli were presented in a set order. The selection of stimuli included three pieces that have a very similar beginning (with the same amount of ambiguity in their jump-phrases). In theory, the third time this beginning is heard, there is already a similar fragment in the memory store. To separate the effect of memory from the effect of the set order will pose a challenge. Bearing the above in mind, we will set about analysing the data with great caution.

Expectations

In the experiment described, subjects with varying degrees of musical experience were asked to parse simple melodies. We expected the variance in assigned boundary points to be smallest for the subgroup with formal musical training (greater degree of consensus), and highest for the musically inexperienced subgroup (lesser degree of consensus). Furthermore, we expected the musically inexperienced group to show fewer exceptions to the Gestalt principles of perception, as might be seen from investigation of jump-phrases in the test stimuli. This expectation is based on the assumption the grouping preferences of musically experienced subjects will be based on more similar constructs, and that musically inexperienced subjects would be more inclined to adhere to general aspects of perception and less sensitive to grouping phenomena that we consider specific for musical material. We expected the moderately experienced group to show the transition between the inexperienced and the highly experienced group.

Method

Subjects

Thirty-three subjects were tested, of which three were excluded because they stated they did not have a background of western tonal music. All subjects, 14 women and 16 men, with an average age of 35,9 years (with standard deviation 12,2 years) were grouped into three subgroups according to their level of musical experience. The first group had virtually no musical schooling. None currently played an instrument or had had any sort of music lessons in the past 5 years. Subjects in the second group all currently played an instrument, or were in a profession that involved music (for instance composer or studio technician). The third group consisted of subjects who had all received formal musical training. Group variables are presented in table 1. Participation took place on a voluntary basis.

Material

Test stimuli were selected from the Essen Folksong Collection. To rule out an effect of complexity of the melody, eight songs were selected from the 213 available children's songs, all of which contain a so-called jump-phrase, as well as phrases that are explained by the Gestalt principle of parallelism as a control. Also, two songs were selected from the Essen database that have a slightly more complex melody. All test stimuli are presented in Appendix B, coded according to method used for the Essen Folksong Collection. The melodies were unknown to all subjects. Some minor changes were introduced so as to increase the number of ambiguous phrases. None had any effect on the overall structure of the melody.

The tempi of the songs were set on 100 bpm for 4/4 and 3/4 measure pieces, and at 50 bpm for 2/4 measure pieces. One song, set in 4/2 measures, was set at 200 bpm. This was done so that faster songs are not harder to parse and the perceived tempo is the same for every piece. All songs were presented mostly within the central pitch octave, as only the key is specified in the coding of the folksongs and not the position of the song.

Procedure

Every session started by explaining the concept of a phrase through a standardised information sheet (Appendix C, in Dutch). An example stimulus on which the subject could practice segmenting a melody by pressing keys on a keyboard. Then, a test set of 10 melodies was presented in a set order. The subjects heard the stimulus tunes through a set of headphones (Monacor MD 5500). They were asked to perform a computer based task, in which the stimuli are presented as an audio file in Cubasis VST, and visualised as a timeline with a moving cursor. By pressing a key subjects indicated where they feel the phrase boundaries should be placed, this resulted in a mark below the timeline. A maximum of three trials was offered per stimulus, in which the subject had the possibility to correct or change an earlier segmentation.

Afterwards, the subjects were asked to fill in a short questionnaire in which information on musical training, cultural background and a few other personal traits was obtained.

Data-analysis

The results of the questionnaire were analysed for information on the number of phrases which were assigned and trials subjects needed to do so, along with several personal variables such as musical experience, and familiarity with the presented melodies.

As there are two specific questions about the segmentation process, one of universality and one of exceptions to the Gestalt principles, the data have been analysed in two ways.

To get an idea of the consensus in the segmenting process, both for the entire group and for the subgroups, every possible boundary point in the data set was reviewed to see how many subjects had assigned a phrase boundary there. A number of critical points was identified, namely where, within a space of three possible boundary points, more than half of the subjects (5 per subgroup, 15 for the total group) had assigned a phrase boundary. The variance of these critical points was calculated and averaged.

To assess the number of exceptions to the Gestalt Principles, the subjects' parses were compared with the ambiguous parses made by the Essen group. Boundary points were rated either as 'same', 'before', 'after' or 'none', thus discriminating assigned boundaries that either separate adjacent same notes into two phrases or keep them together. According to the ambiguity of the phrase boundary, boundary points were rated as jump-phrases or non-ambiguous. The boundary points that have the same amount of ambiguity were averaged in analysing the responses to these phrases, according to melodic ambiguity, temporal ambiguity and ambiguity in parallelism.

Qualitative analysis concentrated on within-subject consistency and on reasons for placing unexpected phrase boundaries.

Results

Table 1 shows several of the subjects' personal traits, as well as the number of phrases they assigned and the number of trials needed to do so. The musically inexperienced group is referred to as 'low', the moderately experienced group as 'moderate', the highly experienced group as 'high'.

Table 1:Personal traits of the total group and subgroups of subjects

	n	sex		age		music lessons (yrs)		number of phrases per tune		number of trials per tune	
		m	f	M	SD	M	SD	M	SD	M	SD
total	30	16	14	35,93	12,19	8,48	10,16	7,59	3,44	1,96	0,66
low	10	4	6	29,50	6,19	0,80	1,17	8,64	4,54	2,11	0,68
moderate	10	6	4	36,70	10,92	7,15	3,25	7,57	3,04	1,85	0,67
high	10	6	4	41,60	15,55	17,50	12,77	6,56	1,93	1,91	0,71

According to the results of the questionnaire, 16 out of 30 subjects reported that the melodies seemed familiar, none reported they positively knew them. 8 out of 30 subjects reported that in assigning phrases they had taken possible lyrics into account (even though they were not present in the stimuli). They all rated the task as relatively easy to perform (mean 3,7 on a 5-point scale).

To assess universality of the assigned phrase boundaries, critical points were identified where in three consecutive boundary placements more than half of the subjects (>5 per subgroup, >15 for the entire group) had assigned a phrase boundary. The amount of critical points that was identified varies per subgroup. For the 'low' group, 61 critical points were identified, for the 'moderate' group this was 58, for the 'high' group this was 50. For the total group the number of critical points was 56. The variance for each critical point was averaged over the total group or subgroup. The results are displayed in Table 2, along with the results of a one-way ANOVA, comparing the different group means.

Table 2:Variance per boundary point per subgroup, p-values for different combinations of means

	critical points	mean variance	standard deviation	*p*	low	moderate	high
total	56	0,291	0,187		0,795	0,318	0,084
low	61	0,300	0,192		x	0,208	0,048
moderate	58	0,254	0,201			x	0,478
high	50	0,227	0,185				x

As displayed in Table 2, the mean variance of the critical points assigned by the experienced group differs significantly from that of the total group ($F(1,104)=3,04$, $p<0.10$) and even more so from the variance of the critical points assigned by the 'low' group ($F(1,109)=4,01$, $p<0,05$). The other means did not differ significantly.
A visual representation of the variance in the parsing process can be seen in Appendix D, showing histograms for every possible boundary point.

To assess the number of exceptions to the Gestalt principles that were seen, a number of phrase boundaries that showed the largest jump were selected. Phrase boundaries that could be explained by parallelism or temporal spacing were excluded. Figure 3 shows how the different subgroups parsed these ambiguous phrases.
The 'low' group showed exceptions to the Gestalt principles in a third (34,5%) of the ambiguous phrase boundaries (they assigned approximately twice as many non-violating boundaries as they assigned violating boundaries), whereas both the 'moderate' and the 'high' group showed these exceptions for about half (resp. 48% and 49,3%) of the ambiguous phrase boundaries.

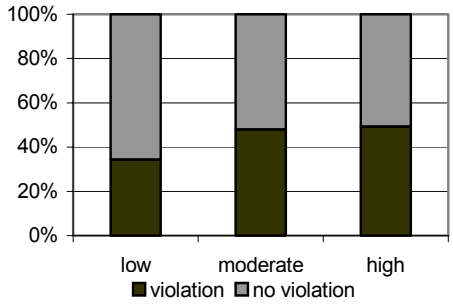


Figure 3: Violation of the Gestalt principles in ambiguous phrase boundaries

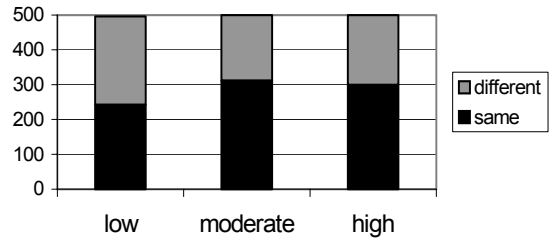


Figure 4: Same/different phrase boundaries of total subject group as compared with the Essen group

When all the phrases made by the Essen University group were compared with the subject groups, about a three-fifths of 50 phrase boundaries were assigned differently by the 'moderate' and 'high' group (resp. 61,4% and 59,8%), the 'low' group assigned almost half (48,0%) of the phrase boundaries differently, as shown in Figure 4. Where minor changes were made in the stimuli, boundaries were taken out of the comparison.

The possible effect of a repeated musical fragment is displayed in Figure 5. Three extremely similar fragments were presented in the course of the testset. The figure shows that the third time it was heard, more subjects assigned the phrase boundary consistent with the Essen group phrasing. Though the effect was modest, the 'low' group showed the strongest effect.

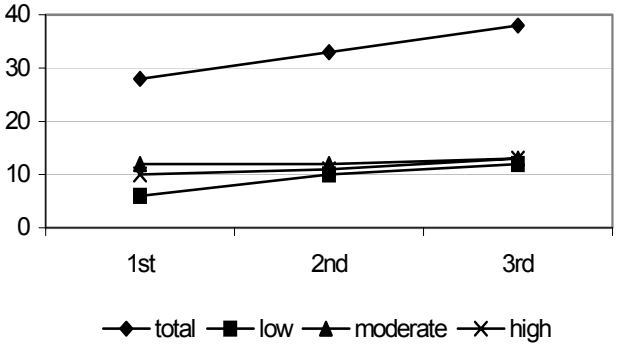


Figure 5: Amount of boundaries assigned same as Essen group over repeated presentation of similar segments

Discussion of the results

Universality

As can be seen in Table 1, the 'low' group assigned more phrases to the tunes than the 'high' group, with the 'moderate' group showing a transition between the two extremes.
Unlike in other experiments of this kind (e.g. Deliège 1987), subjects could assign as many phrase boundaries as they wanted. Specifically, they were asked to assign the shortest phrases they thought reasonably possible. This resulted in considerable differences in number of phrases per tune per subject (seen in the standard deviation for the number of assigned phrases in Table 1), which limits the possibilities for data-analysis.
As expected, the highly experienced group showed more consistency in their parsing process, as can be seen from the lower standard deviation in their number of parses and their lower variance in their segmentation patterns in comparison with the 'low' group. The transition from inexperienced to highly experienced does not reach significance when comparing the variance of the assigned critical points, although a weak effect may be seen.

Exceptions to the Gestalt principles

In this experiment, there does not seem to be a difference between moderately experienced and highly experienced listeners when it comes to showing exceptions to the Gestalt principles. Both these groups assign approximately half of the ambiguous jump-phrases in accordance, and half in violation with the Gestalt principles. This shows that for the more

experienced listener, these phrase transitions are truly ambiguous. For the inexperienced group however, only about a third of the responses were in violation with the Gestalt principles. This confirms our expectation that less musical experience leads to fewer exceptions to the Gestalt principles.

When all the phrases that were assigned by the Essen group were compared to this subject group, again the 'moderate' and 'high' group showed similar responses, rating about three-fifths of the phrase boundaries the same. The 'low' group assigned a little less than half of the same boundaries as the Essen group did. This implies that the coding of the folksongs in the Essen Folksong Collection, done by musically trained people, reflects a consensus of about 60% of the musically trained population, as well as moderately experienced listeners. At face value, this does not seem to be a very strong consensus. However, when reviewing the critical points, out of 50 reviewed phrase-boundaries, the moderate group omitted two of the Essen boundaries and the highly experienced group omitted four. All other phrase boundaries assigned by the Essen group was placed in one of the critical points where more than half of the respective groups had assigned a phrase boundary. The phrase boundaries that were added by the subject group are discussed in the qualitative analysis. The qualitative interpretation and discussion of the generalizability of the Essen group segmentation falls outside the scope of the questions addressed here.

Memory-based parsing

When three consecutive presentations of very similar fragments of music were compared, it seems that by the third time, especially inexperienced listeners improve in their perception of internal structure, implying that repeated listening to an excerpt results in a improved depth of processing, instead of repeating (and strengthening) the initial response.

This modest effect is interesting, but cannot be considered a conclusive finding. Even though the increase in perceived structure for the inexperienced group might be attributed to memory-based processes, the set-up of this experiment doesn't allow for a clear relation of causality. The effect that was found could also be caused by practice, or some effect of the order in which the tunes were set.

Qualitative analysis

There is a confounding factor in this analysis that is not revealed through the difference in means of variance of critical points as they were calculated. The response pattern of one subject is lost in the combined score of the subgroup. Qualitative analysis however yields information on the consistency with which subjects chose to violate Gestalt principles or not. This within-subject consistency was much higher for highly experienced listeners. Even though they might not share their opinions on what the correct way is to segment a tune, they were more consistent (and thus probably more aware of) the choices they made. This consistency can be seen in, for instance, choosing to either always include or always exclude the pick-up note in placing the phrase boundaries.

Another aspect of the data that does not become clear in the above stated analyses, is the reason for the placements of the critical points in the melodies. On closer observation, it appears that the 'low' group made boundary judgements for 26 critical points in places that

weren't accounted for by the Essen group phrasing. This includes both differently placed boundary points as well as omissions of expected boundary points. For the 'moderate' and the 'high' group, this number was much lower (resp. 10 and 9 points). The reason why these points were placed is of interest in achieving a theoretical concept of what constitutes the perception of a boundary point. For the 'low' group, the majority of unexpected boundary points was caused by a temporal jump (a long duration of a note between shorter ones) or rhythmic factors (for instance, if the beginning of a measure constitutes the beginning of a new phrase). Pitch jumps contributed to a lesser amount of unexpected boundary points for the 'low' group, but nevertheless are an important factor in structural grouping. For the 'moderate' and 'high' groups, temporal and pitch jumps each accounted for about half of the unexpected boundary points. It seems that inexperienced listeners are more sensitive to temporal factors, whereas experienced listeners are not as easily distracted from the melody line.

The effect of the more complex tunes was clearly visible, for these melodies the variance was substantially greater for all subgroups. However, the test set (two songs) was too small to provide us with any conclusive findings.

Confounders

Before we set to interpreting these results, a number of possible confounding variables must be mentioned. As the stimuli were presented in a set order, the results must be interpreted with caution. Even though no direct effect on consensus or interpretation of ambiguity can be seen over the course of the stimuli, an effect of the set order in which the tunes were presented cannot be ruled out. Also, one has to keep in mind that relatively simple, monophonic, ambiguous melodies were selected on purpose, which means that the results obtained may well be generalizable to ambiguous melodies only, however simple. If only non-ambiguous tunes had been selected, the consensus we see now would no doubt be much greater. For a visual representation of a highly ambiguous tune compared to a tune with almost no ambiguity, see songs A0214 in comparison to song K0121B in Appendix D.

As the subjects were not matched for age, there is a possibility that the perceived differences between groups are in some way cohort-effects, and a culturally (or rather, historically) based differences are causing the groups to segment melodies in a different manner. However, considering the variance within each group, this does not seem plausible.

Theoretical Perspectives

To summarize, we found that highly experienced listeners show significantly less variance in their segmentation than inexperienced listeners. Also, both highly and moderately experienced listeners show a significantly larger number of exceptions to the Gestalt principles in their manner of parsing than inexperienced listeners. The results of the current experiment have several implications for the theories discussed earlier.

Generative Theory of Tonal Music

As discussed earlier, Lerdahl & Jackendoff's (1983) GTTM is based on two idealizations. The first one is that of the 'ideal listener'. However, the results of this experiment imply that the more experienced the listener is, the more exceptions to the preference rules stated in the GTTM will be apparent in the parsing process. The authors state that their rules reflect the innate understanding of a musical piece. The data suggest that the more experienced listener is able to ignore these innate mechanisms and perceive more higher-order structure in music than listeners with little to no experience. The GTTM preference rules might indeed be described more accurately as preferences of inexperienced listeners, whereas in the current experiment, experienced listeners adhere to the stated preference rules in only about 50% of the presented ambiguous phrase boundaries. Clearly the amount of musical experience is an important factor in the generalizability of the GTTM.

In previous research, Deliège (1987) asked musicians and non-musicians to place segment boundaries between a line of dots. She found that both groups did this according to the preference rules stated by the GTTM. However, this was a forced-choice paradigm, only one boundary could be placed per musical excerpt. In this manner the salience of the different preference rules could be assessed, finding that a change of timbre is one of the most salient cues for boundary placement. To assess grouping preferences in a context closer to real-time listening processes, in this experiment, no such choices had to be made. It appears that in real-time listening without restrictions on boundary placement, the preference rules as stated in the GTTM might be less valid. On the other hand, the idealization of 'final understanding', the second of the two idealizations, implies that the GTTM is not concerned with real-time listening processes. The qualitative observation of the changes made in parses on second or third listening suggest the opposite. Whereas the first impression of a piece may be mediated by Gestalt-oriented grouping rules, the structure perceived after the entire melody is heard seems to overrule this first impression. This is in accordance with the notion that more musical experience brings more perceived structure to a piece, even though this does not appear to correspond with GTTM's preference rules, which are, after all, based upon the Gestalt principles.

Gestalt theory

According to the above stated description of Gestalt theory, the organizational principles of perception suggest a predisposition for perceiving stimuli in one way rather than another. We may accept such a proposition at the same time as accepting that such predispositions may also be generated by other factors. Similarly, we may accept the Gestalt principles while at the same time regarding other aspects of perception as being learned and culturally variable rather than innate. The Gestalt principles can be seen as reinforcing the notion that the world is not simply and objectively 'out there' but is constructed in the process of perception. This process includes several levels of processing, and the data from this experiment show that in the case of music, the perception of higher-order structure has effect on basic grouping mechanisms. This higher-order structure is most probably influenced by the culturally variable knowledge of musical syntax, including, for instance, the way text is normally set to

music in a certain culture. It is assumed that text is the most salient cue in grouping or phrasing vocal music.

Music

Cues that turn out to be salient for basic grouping in melodies are temporal structures and pitch structures. Palmer and Krumhansl (1987a, 1987b) found these two aspects to be independent of each other, but to have an additive effect on phrase judgement. They found this effect for musicians of varying experience. In the context of the experiment presented here, it may be assumed their entire subject group would qualify as either moderately or highly experienced. The results we have discussed imply that indeed the amount of experience does not mediate this mechanism as much as the presence of any experience. Deliège and Mélen (1997) argue that though musicians and non-musicians use identical means to process music, differences between the two groups can be largely attributed to the efficiency of their use of those processes. Cues collected by musicians tend to be richer, broader and more relevant than cues picked up by non-musicians. The differences seen in this experiment seem to imply that the perception of higher order structure of melodies is most likely an automatic process for highly experienced listeners, whereas this is probably not the case for inexperienced listeners. This would also explain the preference of the inexperienced listener for assigning phrase boundaries that do not violate the Gestalt principles of organization analogous to visual organization.

Another possibility is that a musical factor, of which the experienced listener is implicitly more aware, governs the perceived structure of melodies, and Gestalt principles are simply overruled by this factor. Huron (1996), also working with the Essen database, has identified a disproportionate number of phrases that exhibit an arch-shaped pitch contour. Even though this might be an artefact of the Essen parsers, it might also reflect a general phenomenon in folk melodies that hasn't been accounted for in music theory. If this is the case, the ambiguous phrase boundaries used as test stimuli in this experiment would not be as ambiguous as assumed. However, the proportion of musically experienced subjects that chose to place their phrase boundaries differently doesn't support this idea.

In conclusion, the results of the experiment reported here imply that musicians are more consciously aware of the choices they make in assigning phrase boundaries, probably caused by an automatic, memory-based perception of melodic structure. The absence of this automatic perception of structure in inexperienced listeners results in more variance in the parsing process and fewer exceptions to the Gestalt principles of structural organization. It appears that the Gestalt principles in melodic perception are not as generally applicable as they are in visual perception, due to the salience of perceived structure (overruling the basic Gestalt principles) which seems to be automatic for both moderate and high levels of musical experience.

References

Bod, R. (2001). Memory-based models of melodic analysis: Challenging the Gestalt Principles. Journal of New Music Research, 31, 27-37

Deliège, I. & Mélen, M. (1997). Cue abstraction in the representation of musical form. In Deliège, I. & Sloboda, J. (Eds.)(1997). Perception and Cognition of Music. Hove: Psychology Press

Deliège, I. (1987). Grouping Conditions in Listening to Music: An Approach to Lerdahl and Jackendoff's Grouping Preference Rules. Music Perception, 4, 325-360.

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. Perception and Psychophysics, 28, 381-389

Huron, D. (1996).The melodic arch in Western folksongs. Computing in Musicology,10, 2-23

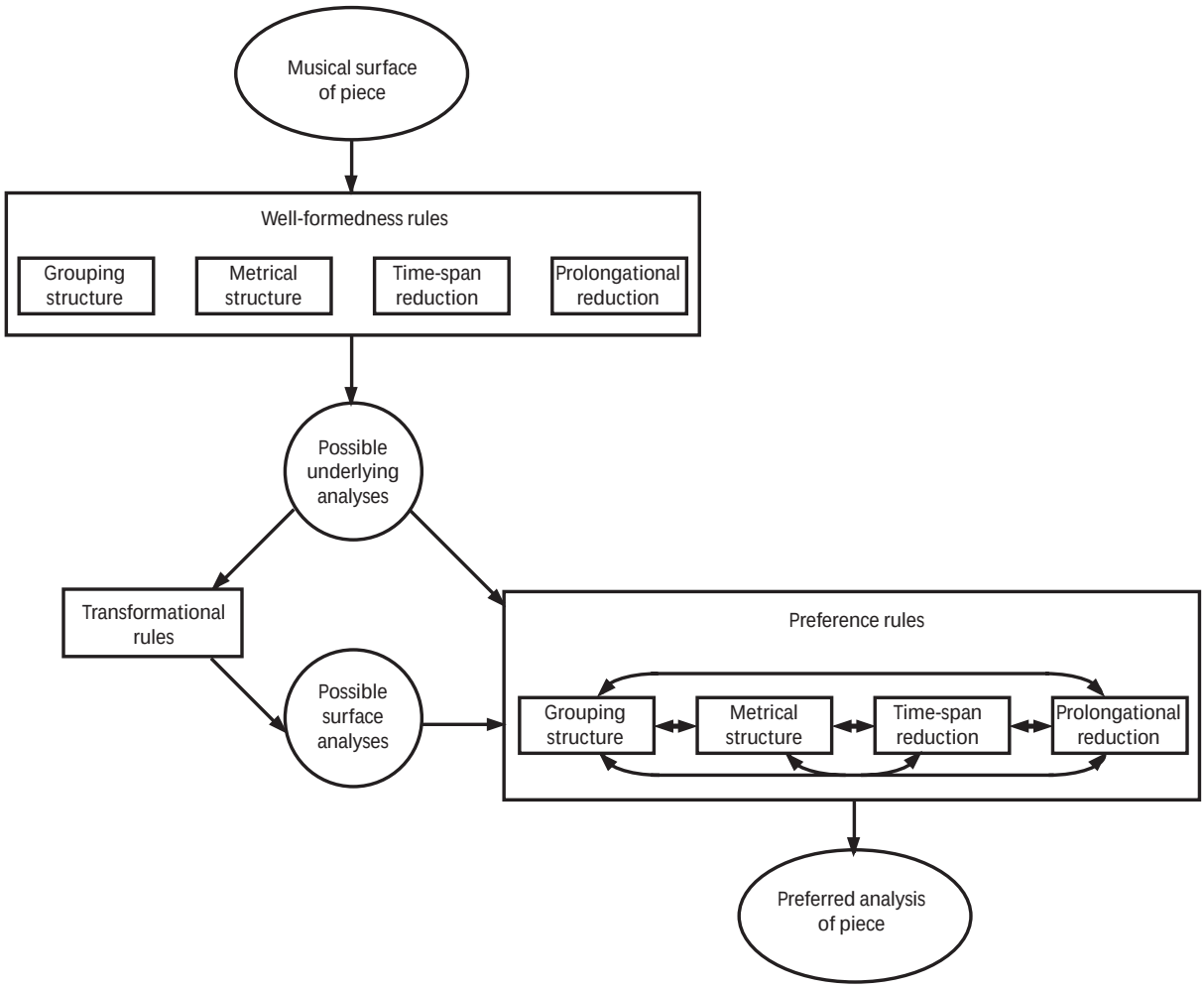
Lerdahl, F. & R. Jackendoff (1983). A Generative Theory of Tonal Music. The MIT Press, Cambridge.

Palmer, C. & Krumhansl, C.L. (1987a). Independent temporal and pitch structures in determination of musical phrases. Journal of Experimental Psychology: Human Perception and Performance, 13, 116-126

Palmer, C. & Krumhansl, C.L. (1987b). Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. Perception and Psychophysics, 41, 505-518

Sloboda, J.A. (1985).The musical mind: The cognitive psychology of music. Clarendon Press, Oxford.

Appendix A:
A schematic overview of Lerdahl& Jackendoff's Generative Theory of Tonal Music(1983)



Appendix B
Teststimuli, in the ESAC

1- SEGG DOCH MUSEKAETTKEN
KEY[K0669 08 Bb 3/4]
MEL[3_3_33 2_2_2_ 1_-7_1_ -5__
-5_ 3_3_3_ 2_2_2_ 1_-7_1_ -5__
3_ 1_-7_1_ 2__
3_ 1_-7_1_ 2__
5_ 3_2_1_ 2_1_-7_ 1__-7_ 1__0_ //] >>

-2- PRIPE NINNE SAUSE
KEY[K0059 16 F 2/4]
MEL[3_3_2_2_ 1__-5_
-5_ 3_3_2_2_ 1__-5_
-5_ 3_3_3_5_ 5_4_4_
3_ 2_2_2_4_ 4_3_3_
1_ 3_3_2_22 1__-5__
3_3_2_2_ 1__-5_0_ //] >>

-3- SO FAHREN DIE DAMEN,
SO REITEN DIE HERREN
KEY[K0352 08 G 3/4]
MEL[-5_ 1_.23_ 2_-5_
-5_ 2_.34_ 3_1_
-5_ -5_-5_-5_ -5_-5_
-5_ -5_-5_-5_ -5_-5_
-5_ 1_1_1_ 1_1_
-5_ 1_1_1_ 1_1_0_
5_-5_-5_ -5_-5_-5_
5_-5_-5_ -5_-5_ //] >>

-4- DER BESTE FREUND
IST IN DEM HIMMEL
KEY[B0752 08 F 3/4]
MEL[1_ 5__5_ 5_4_3_ 5_4_3_ 2_1_
-5_ 4_3_2_ 3__3_ 5_4_3_ 2__
1_ 5__5_ 5_4_3_ 5_4_3_ 2_1_
-5_ 4_3_2_ 3__3_ 5_4_3_ 2__
2_ 2__2_ 3__3_ 4_3_4_ 5__
5_ +1_7_6_ 5_.654 3__2_ 1__ //] >>

-5- SCHLAF KINDLEIN FESTE
KEY[K0029 08 F 2/4]
MEL[3_22 1_-5
-5 3322 1_-5
1 3335 432
1 3335 432_
3_22 1_-5_ //] >>

-6- MUTSCHEKUEHCHEN
VON HALBERSTADT
KEY[K0121B 16 F 2/4]
MEL[5_.43_2_ 1_1_-5_0_
5_.43_2_ 1_1_-5_
-5_ 1_2_3_5_ 5_4_2_0_
1_2_3_5_ 5_4_2_0_
5_.43_1_ 4_6_5__
4_2_3_1_ 2_-7_1_0_ //] >>

-7- DORT OBEN AUF DEM BERGE
KEY[K2223 16 F 2/4]
MEL[-5_ 1_1_1_1_ 1__-5_
-5_ 3_3_3_3_ 3__0_
1_ 5_5_5_5_ 5_3_1_
1_ 2_2_-5_-5_ 1__0_
55 5_4_3_55 5_4_3_
33 2_2_2_2_ 1__0_ //] >>

-8- [Absage]
KEY[A0214 04 G 4/2]
MEL[1_ 1_1_1_1_ -7b_-7b_-5_
-5_ 3b_.3b4_3b_ 2__1_
1_ 1_1_1_1_ -7b_-7b_-5_
-5_ 3b_.3b4_3b_ 2__1_
1_ 3b_.45_3b_ 4__1__
4_.43b_21 -7b_.12_
3b_ 4_1_3b_2_ 1__. //] >>

-9- O DU KLEINER KLADRIAN
KEY[K0641 08 G 2/4]
MEL[11-7-5 11-5
-5 11-72 1_-50
11-7-5 222_
11-72 1_-5
-5 11-5-5 11-5_
11-7-5 222
2 11-72 1_-50 //] >>

-10- RURU RINNEKEN
KEY[K0690 16 F 2/4]
MEL[3__2__ 1_1_-5_
-5_ 3_3_2_2_ 1_1_-5_
-5_ 1_2_3_1_ 4__2_
1_ -7_1_2_-5_ 3__1_
3_ 1__-5_3_ 1_1_-5_3_ 1__-5_
-5_ 1_2_3_1_ 4_3_223_ 1__1_0_ //] >>

Appendix C: Standardised Informationsheet, in Dutch

Informatieblad onderzoek muzikale segmentatie

Als we een melodie horen delen we die meestal automatisch in stukken. Soms is er een duidelijk gevoel van ‘vraag’ en ‘antwoord’, met als simpel voorbeeld:

*Altijd is Kortjakje ziek (?), Midden in de week maar ’s zondags niet (.)

Hierbij is sprake van twee ‘zinnetjes’ of frasen. Soms is er niet eens echt sprake van vraag of antwoord, maar deelt het liedje zich toch duidelijk in stukken op. Bijvoorbeeld:

*Sinterklaas kapoentje/gooi wat in m’n schoentje/gooi wat in m’n laarsje/dank je sinterklaasje

Deze stukken zijn erg kort, bij iets minder simpele melodieën zullen de frasen iets langer worden. Deze langere frasen zijn vaak weer onder te verdelen in kortere frasen. Bijvoorbeeld:

*De herdertjes lagen / bij nachte / zij lagen bij nacht in het veld / ze hielden getrouwe /
de wachte / zij hadden hun schaapjes geteld / etc

Uit dit voorbeeld blijkt ook dat een frase niet altijd op de tel hoeft te beginnen, hoe de muzikale zinnetjes lopen is onafhankelijk van het ritme van de melodie.
De kortst mogelijke frasen zijn vaak, maar niet altijd de duidelijkste, zo is

*vader jacob, vader jacob

één frase, terwijl je een enkele ‘vader jacob’ ook als frase zou kunnen zien. Dit wordt echter wel erg kort, en is eigenlijk een subfrase. Dit is een klein stukje van de melodie, dat niet goed op zichzelf kan staan.

In het volgende onderzoek zal je gevraagd worden bij het luisteren naar een simpel melodietje door op een toets te drukken aan te geven hoe jij denkt dat de frasen lopen en waar jij vindt dat de grenzen liggen. Bedenk wel, alle noten (en stiltes tussendoor) moeten ergens bij horen! Probeer de frasen voor je eigen gevoel zo kort mogelijk te maken. Door op een toets te drukken geef je aan dat alles vóór dat punt bij het vorige deel hoorde, en erna bij een nieuw deel. Er is geen tekst beschikbaar.

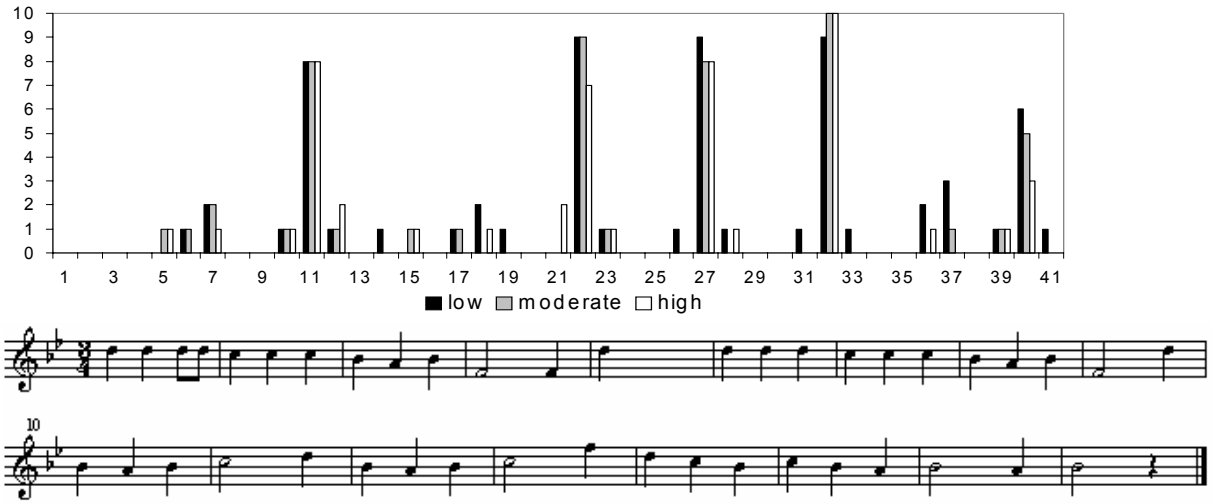
Het eerste melodietje is een oefenvoorbeeld, en zodra je onder de knie hebt hoe je een liedje in frases verdeelt gaan we verder met het onderzoek zelf. Je mag elk melodietje maximaal 3 keer beluisteren en eventueel wijzigingen aanbrengen, je gebruikt hiervoor eerst de ‘1’-toets, dan de ‘2’-toets en dan de ‘3’-toets.

In dit onderzoek wordt gekeken of iedereen dit op dezelfde manier doet, er zijn dus geen ‘foute’ antwoorden. Probeer aan te geven hoe het liedje op jou overkomt, en welke noten volgens jou bij elkaar horen. Het liedje wordt vrij langzaam afgespeeld, zodat je de kans hebt tussen elke twee mogelijke noten een grens aan te geven.

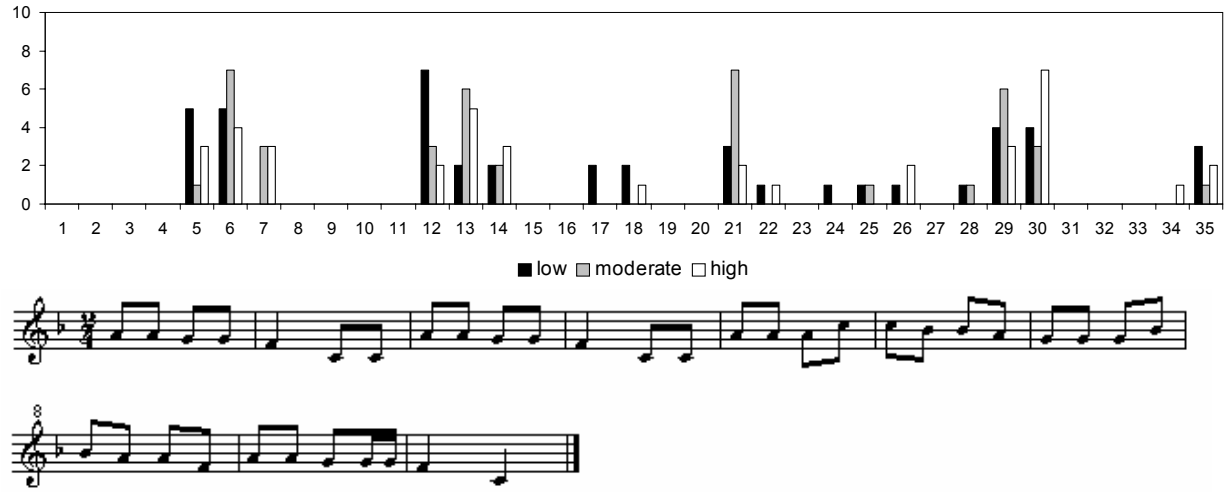
Als het je nog niet helemaal duidelijk is, vraag dan om meer uitleg! Dit kan op elk moment tijdens het onderzoek.

Appendix D: Visual representation of universality for all test stimuli.

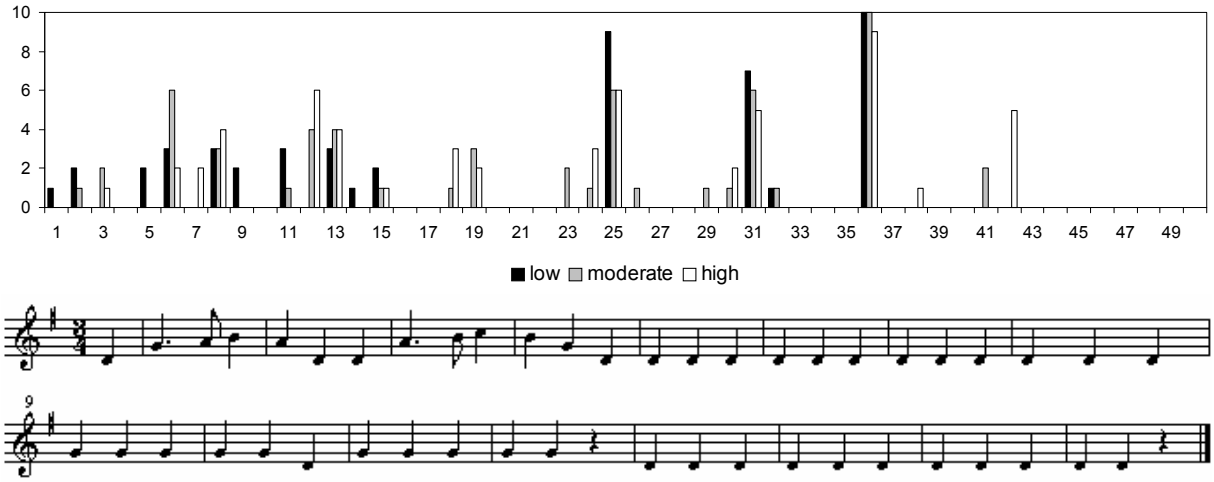
The horizontal axis displays all possible boundary points, the vertical axis the number of subjects that assigned a phrase boundary at that point. Below the graphs a representation of the stimulus is presented in sheet writing. The numbers on the horizontal axis represent the consecutive spaces between the notes, i.e. 1 is the space between the first two notes, 2 is the space between the second and third notes, and so on.



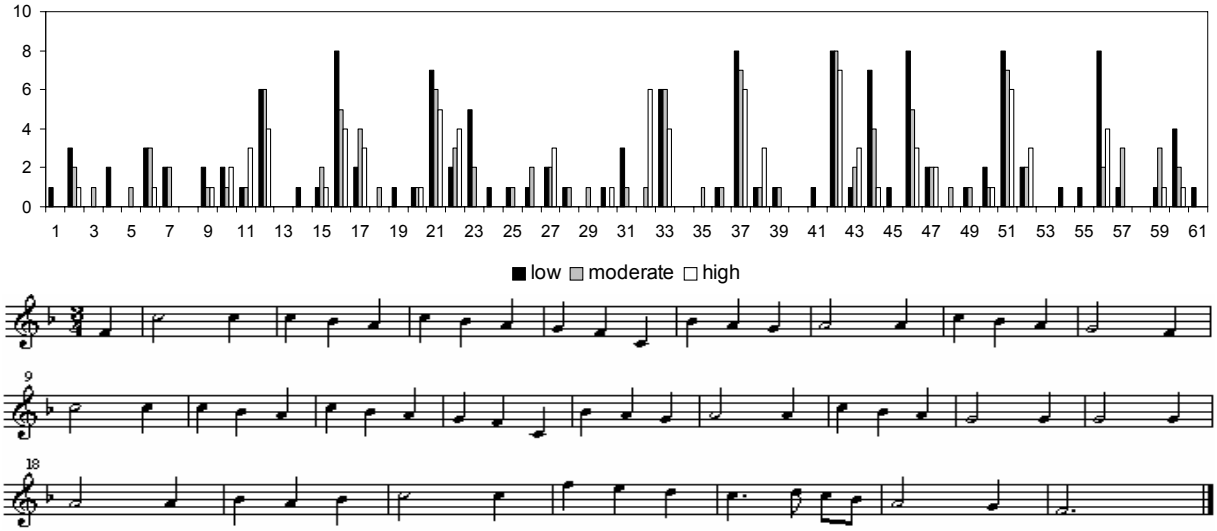
1. K0669



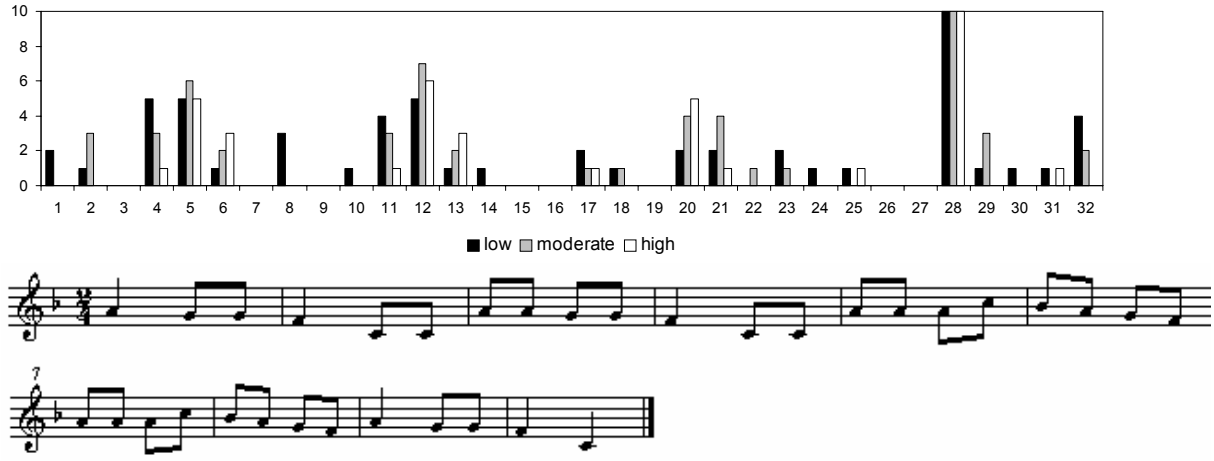
2: K0059



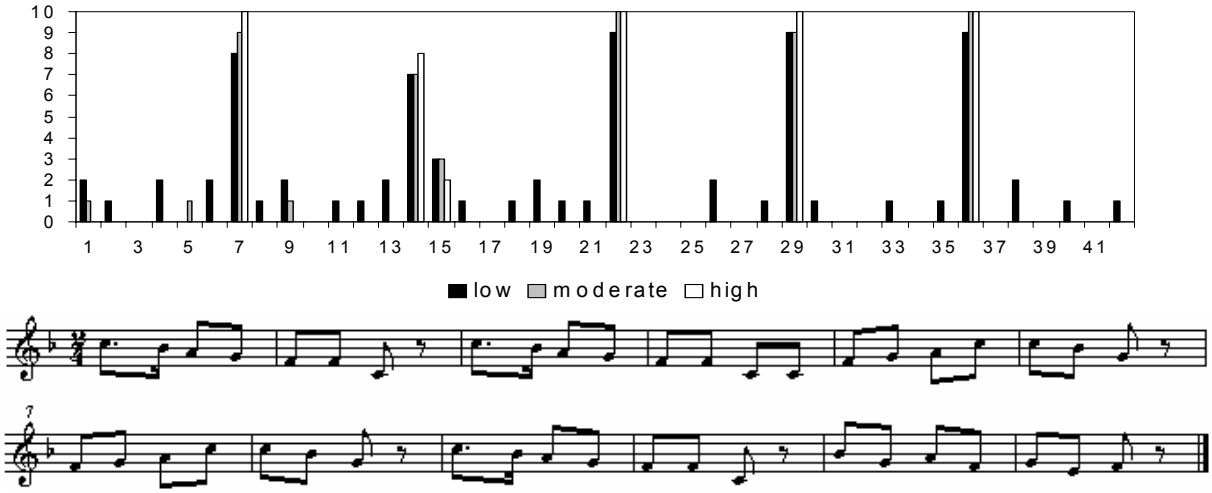
3. K0352



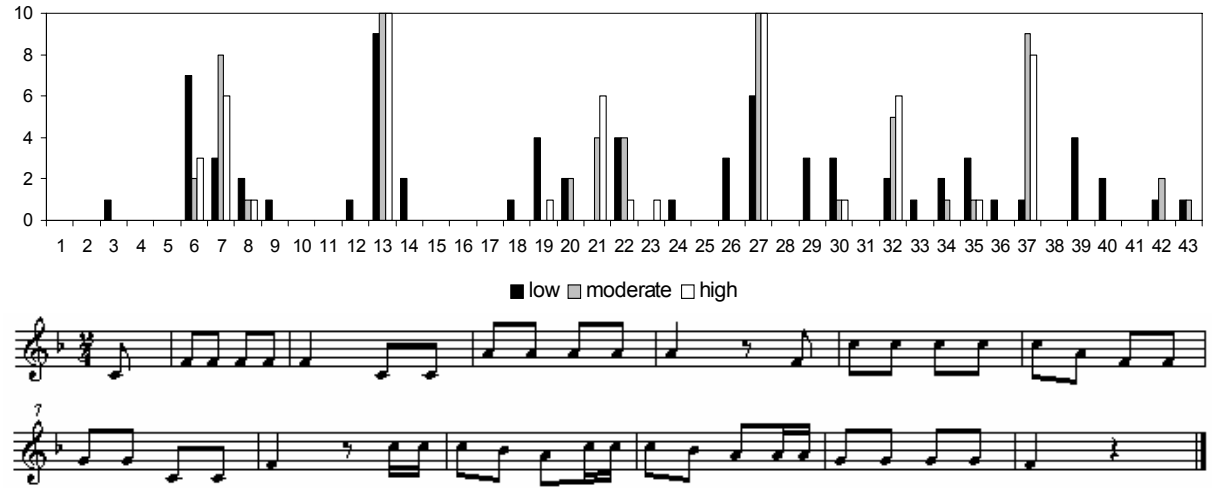
4. B0752



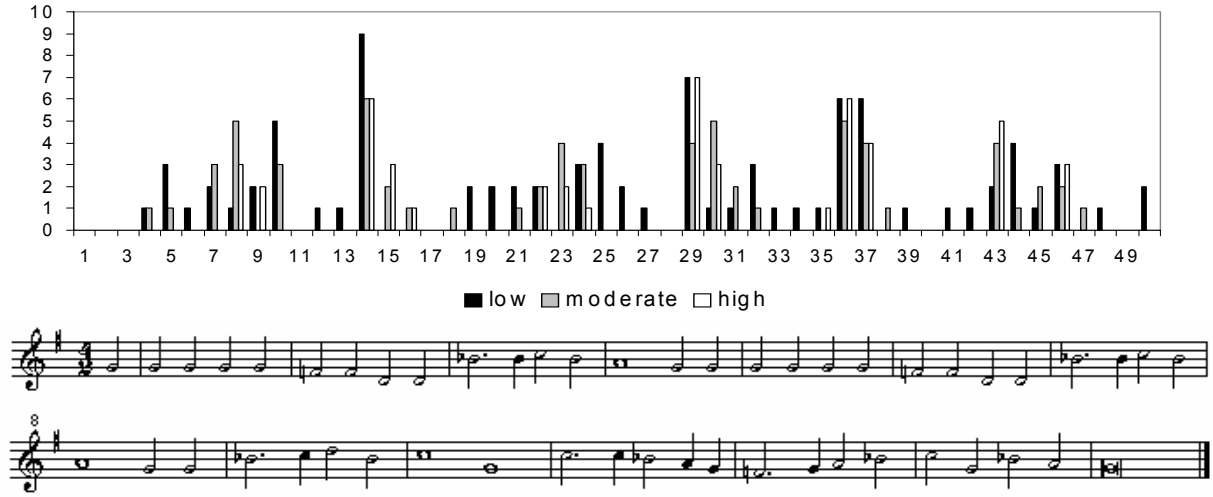
5. K0029



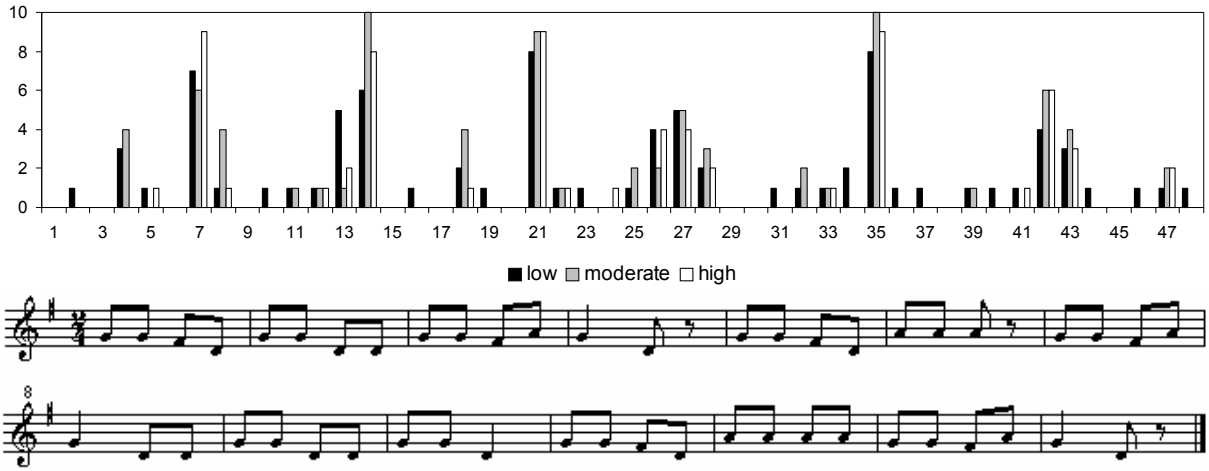
6.K0121B



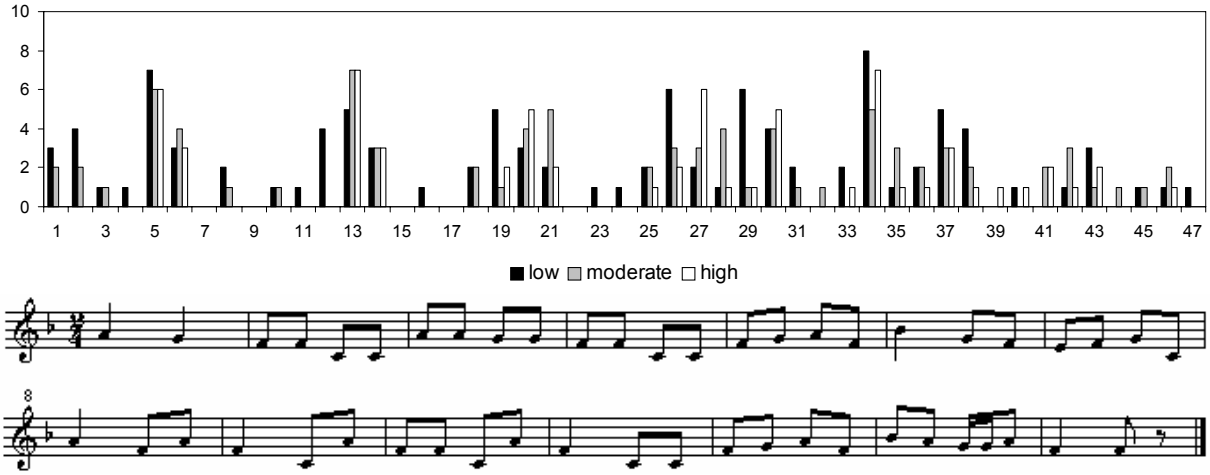
7. K2223



8. A0214



9. K0641



10. K0690