The effects of rhythm and melody on auditory stream segregation

Running title: Streaming by melody and rhythm

Orsolya Szalárdy

Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, H-1519 Budapest, P.O. Box 286, Hungary

Alexandra Bendixen

Auditory Psychophysiology Lab, Department of Psychology, Cluster of Excellence "Hearing4all", European Medical School, Carl von Ossietzky University of Oldenburg, Ammerländer Heerstr. 114-118, D-26129 Oldenburg, Germany

Tamás M. Bőhm

Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, H-1519 Budapest, P.O. Box 286, Hungary

Lucy A. Davies and Susan L. Denham

Cognition Institute and School of Psychology, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK

István Winkler[a)]

Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, H-1519 Budapest, P.O. Box 286, Hungary

[a)] Author to whom correspondence should be addressed. Electronic mail:

winkler.istvan@ttk.mta.hu

Abstract

Whilst many studies have assessed the efficacy of low-level similarity-based cues for auditory stream segregation, much less is known about whether and how the larger-scale structure of sound sequences support stream formation and the choice of sound organization. In two experiments, we investigated the effects of musical melody and rhythm on the segregation of two interleaved tone sequences. The two sets of tones fully overlapped in their pitch ranges, but differed from each other in interaural time and intensity differences. Unbeknownst to the listener, separately, each of the interleaved sequences was created from the notes of a different song. In different experimental conditions, the notes and/or their timing could either follow those of the songs, or they could be scrambled or, in case of timing, set to be isochronous. Listeners were asked to continuously report whether they heard a single coherent sequence (*integrated*) or two concurrent streams (*segregated*). Although temporal overlap between tones from the two streams proved to be the strongest cue for stream segregation, significant effects of tonality and familiarity with the songs were also observed. These results suggest that the regular temporal patterns are utilized as cues in auditory stream segregation and that long-term memory is involved in this process.

PACS numbers: 43.66.Mk, 43.75.Cd, 43.66.Lj

I. INTRODUCTION

In everyday life, the ears continuously receive multiple sounds originating from concurrently active sound sources. In order to adapt to and efficiently interact with our environment, we need to organize these acoustic events into coherent sequences, termed *auditory streams* which are usually associated with separate sound sources (Bregman, 1990; Winkler et al., 2009a). The formation of auditory streams has been investigated within the framework of *auditory scene analysis* (Bregman, 1990; Snyder and Alain, 2007; Winkler et al., 2009a). Several studies have shown that the auditory system forms streams by grouping together sounds with similar acoustic features (see e.g., Moore and Gockel, 2002). Recently, Bendixen et al. (Bendixen et al., 2013; Bendixen et al., 2010) showed that the auditory system can utilize regular temporal patterns for sorting sounds into streams (see also Andreou et al., 2011; Rimmele et al., 2012; Snyder and Weintraub, 2011). However, while the regularities tested by Bendixen and colleagues (Bendixen et al., 2013; Bendixen et al., 2010) extended the intervals during which listeners perceived the tone sequence in terms of two streams, they did not affect the length of the intervals during which listeners perceived the tone sequence as a single integrated stream. This was taken to suggest that proto-objects (groupings of sounds that may appear in perception) may have two separate attributes that affect their competition for dominance. One is their ability to wrest dominance away from the currently dominant (perceived) proto-object. We term this attribute *competitiveness*. A proto-object with high competitiveness will shorten the dominance duration of other proto-objects. Thus the effect of a cue on the competitiveness of a proto-object can be assessed by measuring its effect on the dominance durations of other proto-objects. The other attribute of proto-objects reflects their ability to resist the competing proto-objects and thus to remain dominant. We term this attribute *stability*. A proto-object with high stability will have long dominance durations. Thus the effect of a cue on the stability of a proto-object can be assessed by measuring the dominance durations of that proto-object itself. Switching occurs when the competitiveness value of a currently non-dominant proto-object exceeds the stability value of the currently dominant proto-object. Competitiveness and stability may be two independent attributes of proto-objects. Alternatively, it is possible that a common "activation" attribute (see Mill et al., 2013) is affected differently by certain cues depending on whether the proto-object is dominant or not. Bendixen and colleagues (2010) implicitly suggested the latter by assuming that the regularity of their temporal patterns is only detected when the corresponding proto-object is dominant. As a consequence, the regular temporal patterns tested by Bendixen and colleagues (Bendixen et al., 2013; Bendixen et al., 2010) appeared only to stabilize auditory streams which have already been segregated but not to induce segregation by themselves. In contrast, the similarity-based cues tested so far (separation in pitch, sound source location, and amplitude modulation rate) not only extended the intervals of segregation but also shortened the intervals of integration (Bendixen et al., 2013; Denham et al., 2013; Szalárdy et al., 2013). On this basis, one may distinguish cues that induce percepts by increasing their competitiveness (*percept-inducing cues*) and cues that stabilize percepts by increasing their stability (*percept-stabilizing cues*); allowing also for cues that exert both effects (*percept-inducing/stabilizing cues*). In the current study, we investigated whether melody and rhythm (two of the structural features characterizing human speech and music) can act as cues for auditory stream segregation and if so, whether they stabilize and/or induce auditory streams. Specifically, we tested whether the presence of melodic and rhythmic patterns, hidden separately in two interleaved tone sequences shortens the intervals during which listeners perceive the sequence as a single stream and/or extend the intervals during which two streams are perceived.

Auditory stream segregation is often investigated in the classical *auditory streaming paradigm*, a stimulus configuration that consist of a repeating 'ABA_' pattern of sounds (van

Noorden, 1975), where ‘A’ and ‘B’ denote two sounds differing in some feature(s) and ‘_’ stands for a silent interval equaling the common duration of the two sounds. Depending on the stimulus parameters, this stimulus configuration is usually perceived in one of two different ways: Either all tones form a single sound stream termed the *integrated* percept or two streams are heard concurrently, one consisting only of the ‘A’ and the other only of the ‘B’ sounds (the *segregated* percept). Whereas the classical view suggested that perception settles on one of the two sound organizations (Bregman, 1990; van Noorden, 1975), more recent studies demonstrated that when listeners are exposed to ‘ABA_’ sequences for a few minutes, perception switches between alternative sound organizations (Bendixen et al., 2013; Bendixen et al., 2010; Denham et al., 2010, 2013; Denham and Winkler, 2006; Gutschalk et al., 2005; Pressnitzer and Hupé, 2006; Szalárdy et al., 2013). This phenomenon is termed *perceptual bistability* and it reflects competition between alternative interpretations of the sensory input (Leopold and Logothetis, 1999; Pressnitzer and Hupé, 2006; Winkler et al., 2012).

Such bistable configurations can be used to investigate how various auditory cues affect the balance between the *integrated* and *segregated* perception of the sequence. Based on the view that auditory perceptual bistability reflects continuous competition between alternative sound organizations (Denham and Winkler, 2006; Denham et al., 2013; Winkler et al., 2009a; Winkler et al., 2012), a continuous measurement of the listener’s perception can be used to assess properties of the competition. For the analysis, the continuous record of the listener’s report is segmented into intervals between two perceptual switches (termed *perceptual phase*; i.e., the unbroken interval within which the listener reported the same percept). The mean durations of these perceptual phases, averaged separately for each possible percept, provide important information about the way a given cue affects auditory perceptual organization (Bendixen et al., 2010; Bendixen et al., 2013; Denham et al., 2013). For example, increasing the pitch separation in an ‘ABA_’ sequence prolongs the duration of *segregated* perceptual phases and shortens the duration of *integrated* phases by causing switches back to segregation (Bendixen et al., 2013). The first effect (prolonging phases of the supported percept) can be regarded as stabilizing the percept, whereas the latter (shortening phases of the alternative percept) shows the effect of the cue on wresting dominance away from other organization in order to induce perception of the supported organization. Inserting regular pitch and intensity patterns separately into the ‘A’ and ‘B’ tone sets in an ‘ABA_’ paradigm prolonged the duration of *segregated* phases, but had no effect on the duration of the *integrated* phases, i.e. they did not cause perception to switch to ‘segregation’ (Bendixen et al., 2010; Bendixen et al., 2013). Basing on these results, Bendixen and colleagues (Bendixen et al., 2010; Bendixen et al., 2013) hypothesized that similarity-based cues can both induce and stabilize a perceptual organization (i.e., they are *inducing/stabilizing cues*), cues based on regular temporal patterns can only stabilize a percept (*stabilizing cue*).

Here we investigated two cues, melody and rhythm, that are conceptually similar to the repetitive patterns used by Bendixen et al. (Bendixen et al., 2010; Bendixen et al., 2013; see also Andreou et al. 2011) in that they provide structure for sound sequences. Rhythm refers to the temporal arrangement of the shorter and longer notes and pauses in a sequence, while melody refers to the sequence of pitches carried by the sounds. Some studies have indicated that the roots of rhythm perception are already present at birth (Winkler et al., 2009b). However, so far few studies have directly investigated the effects of rhythmic structure on auditory stream segregation (but see Andreou et al., 2011; French-St George and Bregman, 1989; Rimmele et al., 2012; Rogers and Bregman, 1993). Jones and her colleagues have long argued for an important role of rhythmic structure in auditory stream segregation (Jones, 1976; Jones and Boltz, 1989), suggesting that temporal predictability guides the grouping of sounds through attentive processes (Demany and Semal, 2002; Devergie et al., 2010). Contrasting results were obtained by French-St. George and Bregman (1989), who found that a temporal regularity inserted into an ‘AB’ sequence had no significant effect on auditory stream integration. Similarly, Rogers and Bregman (1993) found that a temporally predictable induction sequence did not increase the segregation of the following ABA_ sequence compared to an unpredictable one. In contrast, two recent studies showed that temporal regularities help auditory stream segregation when listeners are instructed to attend one of the streams, even when regular temporal patterns appear only in the other stream (Andreou et al., 2011; Rimmele et al., 2012). Whether the same is true for unbiased listening instructions (i.e., when the listener is not instructed to attend one of the streams), remains to be tested.

The effects of melodic structure on streaming are also not well understood. Dowling and colleagues (Dowling, 1973; Dowling et al., 1987) showed that when a familiar melody was presented to listeners together with distracting sounds taken from the same pitch range, many listeners were able to separate the melody from the other sounds. A similar study was conducted by Bey and McAdams (2002), who presented listeners with two melodic patterns, each consisting of six tones. One of the patterns was interleaved with distractor tones. Listeners were then asked whether the two melodies were identical or not. Listeners’ recognition performance was higher when the target pattern was presented first without the distractor tones. The authors interpreted this result as showing that sequential stream segregation can be improved by previous knowledge, which is mediated by expectations (for a similar conclusion, see Devergie et al., 2010). However, in these studies, listeners were actively searching for familiar melodies, thus they intentionally attempted to segregate the interleaved sequences as it helped them to detect the target patterns. Several studies (e.g., McDermott et al., 2011) showed that familiarity with some given sounds (either through previous knowledge or exposure to the sounds prior to testing them as part of a mixture) enhances the likelihood of segregating them from other concurrent sounds even without the listener actively searching for the familiar sounds (for a review, see Snyder et al., 2012).

Streaming by melody and rhythm

In the current study, we aimed at testing the effects of rhythm and melody on stream segregation without asking listeners to actively search for a particular rhythm or melody. Thus the task did not suggest them that segregation was preferable over integration. The two cues were separately tested in a 2 x 2 fully crossed design by inserting into sequences of interleaved ‘A’ and ‘B’ tones (an ‘ABAB...’ sequence) a) two different melodies (one for the ‘A’ and the other for the ‘B’ tones), including their rhythmic structure, or b) the two melodies, each delivered at a uniform presentation rate, or c) separate randomized sequences of the notes of the two melodies, delivered with their original rhythm, or d) randomized sequences of the notes delivered at a uniform presentation rate (the baseline condition). In all conditions, ‘A’ and ‘B’ differed by an interaural time and intensity difference (ITD and IID, respectively) that promoted the segregation of the two sets of sounds despite the almost complete overlap between the pitch ranges of the interleaved melodies. This allowed detecting potential *stabilizing cues*, which only exert their effect on streams already segregated on the basis of some *inducing cue*. Separation in virtual source location has been previously found to support auditory stream segregation, although compared with other cues, such as pitch separation, it is not a strong cue (e.g., Bőhm et al., 2013; Denham et al., 2010). If the melodic and/or rhythmic patterns hidden in the sequences are utilized as stream segregation cues, the proportion of segregation should increase in the corresponding experimental conditions compared to the baseline condition. Depending on whether this putative increase in the proportion of segregation is caused by prolonging the duration of segregated percepts alone or (also) by shortening the duration of *integrated* percepts, the percept-inducing and stabilizing qualities of melodic and rhythmic patterns will be characterized.

Finally, we also wished to separate the effects of pattern recognition from those of regularity detection on auditory streaming. Therefore, we manipulated the familiarity of the melodies in the experiment by presenting well-known Hungarian and German songs to Hungarian and

English listeners. If familiarity (pattern recognition) is a necessary prerequisite for utilizing melodic and/or rhythmic cues, then Hungarian listeners should report more segregation for the Hungarian compared to the German melodies, whereas English listeners should show no such bias.

Experiment I investigated the questions detailed above. Experiment II was designed to determine the acoustic cues underlying the effects found in Experiment I.

II. EXPERIMENT I

A. Methods

1. Participants

Twenty-five healthy Hungarian and twenty-one English participants took part in Experiment I (18-26 years of age, average age: 21.68, 10 female in the Hungarian and 18-35 years of age, average age: 21.00, 18 female in the English group). Three Hungarian and two English participants were rejected from the analysis: one of the Hungarian participants had difficulties in performing the task, reporting after the experimental session that he could not distinguish the different percepts; and two didn't report perceptual switches throughout the whole experimental session; in the English group, one participant was rejected as she perceived neither the *integrated, segregated* or *both* percepts for more than 40% of the time, across all conditions, and another was discarded as she only perceived the *segregated* percept, across all conditions.

Written informed consent was obtained from participants after the experimental procedures were explained to them. Participants were screened for intact hearing by audiometry before the start of the experiment. The criteria were that the hearing thresholds should not exceed 30 dB HL within the 250 to 8000 Hz frequency range, separately within each ear, and that threshold differences between the ears should not exceed 10 dB at any frequency. The study was approved by the Ethical Committee of the Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, MTA and the University Human Research Ethics Committee of Plymouth University's Faculty of Science and Technology, where the experiments were conducted. Participants received modest financial compensation in Hungary and Plymouth Psychology Participation Pool points (which count toward course credits) in the UK.

2. Apparatus and stimuli

Participants were presented with sequences of complex tones composed of 8 harmonics (weightings: 0.3, 0.15, 0.3, 0.09, 0.07, 0.05, 0.03 and 0.01; all starting in sine phase; tone duration 200 ms, including 50 ms onset and 50 ms offset ramps) alternating in interaural time difference (ITD, +/-100 micro-seconds) and, congruently, in interaural intensity difference (IID, +/-3dB), thus alternating tones were perceived as left- or right-lateralized. The perceived location difference provided the basis for the participants' task (see *Experimental procedure* section below), and with no other cues, resulted in an approximate balance between the *segregated* and *integrated* perception of the sequence (see Results). The tones arriving from the same virtual source location were the notes of one of 4 different songs (Figure 1); the two interleaved sequences were always derived from different songs. The mean frequencies of the songs were 551.4 Hz (377.5 - 635 Hz, spanning 9 semitones [ST]) for the first Hungarian song (H1), 570.9 Hz (423.8 – 847.6, spanning 12 ST) for H2, 585.2 Hz (475.7 – 712.7 Hz, spanning 7 ST) for the first German song (G1), and 568.7 Hz (G2: 423.8 – 712.7 Hz, spanning 9 ST) for G2. Thus the largest difference between the mean frequencies for any pair of the four melodies was 1.03 ST. The base frequencies of all tones fell within the 377.5 - 847.5 Hz range (spanning 14 ST in the B3-C5 range with H3 set at 400 Hz).

Separately for the left and right-side sequences, the onset-to-onset interval (termed stimulus onset asynchrony: SOA) between consecutive tones was either uniformly 400 ms (Constant-Rhythm conditions) or varied between three discrete values, 200, 400, and 800 ms (Original-Rhythm conditions), depending on the length of the corresponding musical note: eighth note - 200 ms, quarter note - 400 ms, and half note - 800 ms. The mean SOAs for each song were: 387.88 ms for H1, 533.33 ms for H2, 475.68 ms for G1, and 434.29 ms for G2. The onset of the right-side sequence was delayed by 200 ms with respect to the left-side sequence, which resulted in an overall uniform 200-ms SOA in the Constant-Rhythm conditions. Familiarity was manipulated by interleaving either the two Hungarian songs (H-H), the two German songs (G-G), or one of the Hungarian with one of the German songs (H-G). The choice of songs for the H-G combination was balanced across participants. Songs on each side were repeated separately until the length reached the four minute target duration of the stimulus blocks. In the Original-Melody conditions, the sequence of tone pitches followed the order of the notes in the corresponding melody. In the Random-Melody conditions, the order of the notes was separately randomized for each repetition of the melody and for each side. Thus the same rhythm and melody manipulation was always applied to both sides.

Fully crossing the three manipulations resulted in twelve types of stimulus blocks for Experiment I: the three combinations of songs (H-H; G-G; H-G) × two types of rhythm (Constant; Original) × two types of melody (Random; Original). Four of these stimulus blocks were delivered twice to test the effects of familiarization with the task and pattern learning during the experiment: the three conditions with constant rhythm and random melody with the three types of song combinations (H-H, G-G, H-G); and the original rhythm/original melody/G-G condition. These four stimulus blocks were first presented at the beginning of the experiment in the following order: constant rhythm/random melody H-H, G-G, H-G conditions and then the original rhythm/original melody/G-G condition. By comparing the listeners' responses between the first and second presentation of these stimulus blocks, we could 1) determine the effects of familiarization with the task (shown by changes from the first to the second presentation of the constant rhythm/random melody stimulus blocks, because it is highly unlikely that listeners would remember long random pitch patterns for several minutes and thus would be helped by pattern recognition) and 2) test whether possible melody-related effects could be based on encountering the melodies multiple times during the experiment (i.e., once a melody following western musical conventions is segregated, if memorized, it could help listeners to segregate the streams whenever this melody is present in one of the streams: increased segregation on the second presentation of the original rhythm/original melody/G-G condition without corresponding changes in the other three repeated conditions). The order of the remaining twelve stimulus blocks was separately randomized for each listener. Overall, the experimental session consisted of sixteen stimulus blocks of 4 minutes duration, each.

3. Experimental procedure

Participants were seated in a comfortable chair in a sound-attenuated chamber, instructed to listen to the sound sequences and continuously indicate their perception with the help of two (Experiment I in Hungary) or three (Experiment I in the UK) response keys. (The reason for the difference was having different response button devices at the two laboratories.) They were asked to depress one of the keys when perceiving all tones - irrespective of their source location - as part of a single coherent sequence (*integrated* percept). Another response key was to be used when hearing two sound sequences in parallel, one from each side, or only one sequence with a uniform source location (*segregated* percept). If the listener heard one sequence containing tones from both sides and, at the same time, another sequence made up of tones with a uniform source location (*both* percept), then Hungarian participants were to depress both keys at the same time, whereas English participants were instructed to use the third response key. Participants were instructed to release all keys when their percept didn’t match any of the types detailed above (*neither* percept). The instructions emphasized to participants that they should keep the appropriate key(s) depressed as long as they heard the given sound organization and to release the key(s) as soon as the perception changed to some other sound organization. The response key assignment was balanced between participants. The experimenter made sure that participants understood the instructions using auditory illustrations. Depending on the participant, the instruction part lasted for 15-25 minutes. Between consecutive stimulus blocks, participants were allowed to relax for 1-2 minutes and longer breaks were inserted after the eighth stimulus block and whenever the participant needed it. The experiment session lasted for ca. 120 minutes.

After the last stimulus block, the experimenter tested whether each participant recognized any of the four songs presented during the session. Each of the four songs was played once in a randomized order. After the presentation of each song, the participant was asked whether he/she recognized the song from the experiment and, separately, whether he/she was familiar with the song (i.e., heard it before the experiment).

4. Data collection and analysis

The state of the response keys was sampled at 250 Hz (4 ms sampling time) and perceptual phases were extracted from the continuous record. As was defined in Introduction, the term ‘perceptual phase’ refers to the continuous time interval during which the same combination of the response keys was depressed (reporting the same perceptual organization). Perceptual phases shorter than 300 ms were excluded from the analysis as these may represent the participants’ inability to precisely synchronize the press and release of the response keys (Moreno-Bote et al., 2010). The proportion of each type of percept (the percentage of the overall time in which each percept was reported) and the average of the logarithm of the phase durations in milliseconds were calculated separately for each participant, condition, and percept (*integrated*, *segregated*, *both* and *neither*). The proportion of a percept which was not reported in a stimulus block was taken to be 0 for the given stimulus block while the corresponding phase duration was set equal to the mean phase duration of the given percept across all conditions for the given participant. As the proportions of the *both* and *neither* percepts were relatively small during the whole experiment (both: 15.7%, neither: 3.9%) and no hypotheses referred to these percepts, these were not analyzed further.

The effects of learning were tested by comparing the proportions and log-mean perceptual phase durations of the *segregated* and *integrated* percepts between the two presentations of the Constant-Rhythm/Random-Melody conditions using mixed model analyses of variance (ANOVAs) with the factors *Group* (English vs. Hungarian; between-subject factor) × *Presentation* (1st vs. 2nd) × *Song* (H-H vs. G-G vs. G-H). Differences between the first and second presentation of the Original-Rhythm/Original-Melody/G-G conditions were tested using two-tailed, paired-sample Student’s *t*-test. Because this analysis was aimed at testing the effects of learning the task or a melody during the experiment, only effects involving the Presentation factor were followed up by post-hoc tests. Based on the results (see Results, *The effects of learning during the experiment*), for the rest of the analysis, we used the second presentation of these four stimulus blocks.

The effects of melody, rhythm and familiarity with the songs were tested using mixed-model ANOVAs of the following structure, *Group* (English vs. Hungarian; between-subject factor) × *Song* (H-H vs. G-G vs. G-H) × *Rhythm* (constant vs. original) × *Melody* (random vs. original), separately for the proportions and the log-mean phase durations of the *integrated* and *segregated* percepts.

ANOVAs were calculated using the STATISTICA software package. Greenhouse-Geisser correction of sphericity violations was applied where applicable and the ε correction factor is reported. η^2 effect size values are also reported for each significant effect and interaction. Post hoc analyses were conducted using the Tukey HSD test. All significant effects and interactions (α=.05) are reported.

B. Results

1. The effects of learning during the experiment

Comparing the first and second presentation of the Constant-Rhythm/Random-Melody conditions, we found a significant main effect of Song (F(2,78) = 7.349, ε = 0.999, $p < .01$, η^2 = 0.159) as well as significant interactions between Group and Song (F(2,78) = 3.806, ε = 0.999, $p < .05$, η^2 = 0.089) and between Group, Presentation and Song (F(2,78) = 4.216, ε = 0.963, $p < .05$, η^2 = 0.098) for the proportion of the *integrated* percept. The interaction between Group, Presentation, and Song was caused by the significant difference between the first presentation of the Constant-Rhythm/Random-Melody/H-H condition and both presentations of the Constant-Rhythm/Random-Melody/G-H condition (Tukey HSD with df = 78: $p <.05$) in the Hungarian but not in the English participants (Tukey HSD with df = 78: $p >.73$). For the proportion of the *segregated* percept, only a main effect of Presentation was found (F(1,39) = 5.704, $p < .01$, η^2 = 0.128). For the duration of *integrated* phases, a main effect of Song (F(2,78) = 3.193, ε = 0.953, $p < .05$, η^2 = 0.076) and an interaction between Song and Presentation were found (F(2,78) = 6.047, ε = 0.949, $p < .01$, η^2 = 0.134). The interaction was due to the first presentation of the Constant-Rhythm/Random-Melody/G-H condition being significantly different from the first presentation of Constant-Rhythm/Random-Melody/H-H (Tukey HSD with df = 78: $p <.001$) condition and also from the second presentation of the Constant-Rhythm/Random-Melody/G-G condition (Tukey HSD with df = 78: p <.05). No significant effects or interactions were obtained for the duration of the *segregated* percept in either group of participants.

The *t*-tests comparing between the two presentations of the Original-Rhythm/Original-Melody/G-G yielded no significant effects in either group of participants. Thus it appears that whereas we found an effect of familiarity with the task on the listeners' perceptual reports, no effect of encountering a melody multiple times was observed.

2. Proportion of the integrated and segregated percepts

Figure 2 shows the proportions of the integrated and segregated percepts in the two participant groups, separately for each condition; each of the panels on the left column and the top row illustrate the data in a structure compatible with the testing of the effect of the main variables, whereas the lower right panel provides summaries of the three panels by collapsing the Song factor. For the proportion of the segregated percept, a significant three way interaction was found between Group, Song and Melody (F(2,78) = 3.350, ε = 0.966, p < .05, η2 = 0.079; Figure 2 upper left panel). Post-hoc tests revealed that this interaction was due to the proportion of the segregated percept being higher for the Hungarian but not for the English participants in the H-H/original-melody conditions compared to any of the random-melody conditions as well as to the G-G/original-melody condition (Tukey HSD with df = 78: p < .05, at least). The Song and Rhythm factors significantly interacted with each other for the proportion of the integrated percept (F(2,78) = 4.039, ε = 0.711, p < .05, η2 = 0.093; Figure 2 lower left panel). This interaction was the result of the proportion of the integrated percept being lower for the H-H and H-G conditions than for the G-G conditions with the original rhythm (Tukey HSD with df = 78: p < .001, at least), but not with the constant rhythm (Tukey HSD with df = 78: p >=.136). Main effects of Song and Rhythm were found for the proportion of both the integrated and the segregated percept (Song: F(2,78) = 16.951, ε = 0.973, p < .001, η2 = 0.303 and Rhythm: F(1,39) = 45.548, p < .001, η2 = 0.539 for the integrated percept; Song: F(2,78) = 11.002, ε = 0.959, p < .001, η2 = 0.220 and Rhythm: F(1,39) = 58.963, p < .001, η2 = 0.602 for the segregated percept; Figure 2 left panels). These main effects reflected that for the original rhythm and for the H-H and H-G songs, the proportion of the integrated percept was lower and that of the segregated percept was higher compared with the other conditions.

3. Phase duration of the integrated and segregated percepts

Figure 3 shows the log mean phase durations of the integrated and the segregated percepts in the two participant groups, separately for each condition in the same structure as was used for the percept proportions in Figure 2. A significant interaction between Song and Rhythm was found for the duration of segregated percepts (F(2,78) = 6.360, ε = 0.944, p < .01, η2 = 0.140; Figure 3 lower left panel). Post-hoc tests revealed that the three levels of Song didn't differ from each other with the constant rhythm (Tukey HSD with df = 78: p >= .778), but with the original rhythm, significantly longer segregated phase durations were reported for the H-H and H-G than for the G-G Song conditions (Tukey HSD with df = 78: p < .001, all). Main effects of Song, Rhythm, and Melody were found on the log mean durations of both the integrated and the segregated percepts (Song: F(2,78) = 4.651, ε = 0.972, p < .05, η2 = 0.107, Rhythm: F(1,39) = 15.846, p < .001, η2 = 0.289, and Melody: F(1,39) = 8.803, p < .01, η2 = 0.184 for the integrated percept; Song: F(2,78) = 5.660, ε = 0.979, p < .01, η2 = 0.127, Rhythm: F(1,39) = 28.955, p < .001, η2 = 0.426, and Melody: F(1,39) = 4.987, p < .05, η2 = 0.113 for the segregated percept; Figure 3 left panels). The integrated phases were shorter and the segregated phases were longer for the H-H and H-G compared with the G G Song conditions, for the original compared with the constant Rhythm conditions, and for the original compared with the random Melody conditions.

4. Recognition of and familiarity with the songs

Table 1 shows the summary of recognition and familiarity answers of the participants.

Hungarian participants were all familiar with both Hungarian songs and most of them recognized them during the experiment. Further, despite very few of them being familiar with the German songs, many still recognized them from the experiment. None of the English participants were familiar with either the Hungarian or the German songs. All subjects, except for two, recognized at least one melody from the experiment.

C. Discussion of Experiment I

We found that both the melody and the rhythm of the temporal tone patterns introduced separately into the two interleaved sequences as well as familiarity with these patterns affected auditory stream segregation: each type of pattern helped participants to segregate the two interleaved tone sequences compared to the base condition, in which tones were delivered isochronously and in randomized order.

The significant main effects of Melody on the duration of both the *integrated* and the *segregated* perceptual phases suggest that conformity with western musical conventions possibly helped to segregate the streams. Two different effects were observed: compared to the random pitch sequences, 1) the *integrated* phases became shorter, which can be interpreted as an increase in the competitiveness of the *segregated* sound organization, and 2) *segregated* phases became longer, which can be interpreted as an increased stability of segregation. This effect was present in both groups of participants: regardless of whether or not participants knew the songs from before the experiment. This interpretation of the results is supported by results showing that listeners familiar with western musical conventions process the tonality of melodies even without formal musical training (Cuddy, 1991; Trainor and Trehub, 1994). However, there is a potential confound in comparing the original and random Melody conditions: the average pitch difference between successive notes of a melody is lower than that between the same notes delivered in a random order. Smaller pitch steps may have increased the coherence of the streams. Although this may have been an important factor, the fact that most listeners recognized the melodies encountered in the stimulus sequences after the main experimental session (Table 1) suggests that they have probably formed memory traces for these melodic patterns, allowing them to utilize the structure of the original melodies.

Bendixen and colleagues (2010) found that regular pitch patterns separately introduced into two interleaved tone sequences extended the *segregated* phases, but did not shorten the *integrated* ones. These authors (see also Bendixen et al., 2013) suggested that whereas similarity-based cues of auditory stream segregation increase both the competitiveness (cutting short *integrated* phases) and the stability of the *segregated* percept (extending the *segregated* phases), predictability-based cues, such as repetitive pitch patterns only affect the stability of the corresponding sound organization. In contrast, here we found both of these effects for the presence of the original melodies. One possibility is that overlearned rules, such as the conventions of western music (for listeners brought up within this musical context), become *percept-inducing cues* of auditory stream segregation. In this case, one may assume that the brain learns to utilize rules, such as tonality, for evaluating similarity between sounds during the initial grouping processes.

Another possibility is that the detection of tonality biased auditory stream segregation by evoking processes operating outside auditory stream segregation (Winkler et al., 2009a). This alternative is supported by an important difference between the current study and those of Bendixen et al. (Bendixen et al., 2013; Bendixen et al., 2010), namely that whereas Bendixen and colleagues used very short (three/four tone) cycles in both streams, our melodies were considerably longer. Thus, whereas the short repeating patterns could be discovered by low-level processes (see, e.g., Sussman et al., 1998), the current melodies could initially only activate the detection of tonality. Detecting tonality could have alerted higher-order systems either to check whether tonality holds throughout or to actually analyze the melody. These processes may have biased the competition between integration and segregation, favoring the segregated solution in which tonality/melody could be analyzed. Thus the difference between the effects found in the current and in Bendixen et al.'s (Bendixen et al., 2013; Bendixen et al., 2010) studies may be related to whether the cues based on regular temporal structure are processed together with the similarity-based cues of segregation (see the account of Bendixen et al., 2010; cf. Winkler et al., 2009a) or at higher levels of the system with access to long-term learned information (see Koelsch and Siebel, 2005). Alternatively, the results can be interpreted as supporting the view that object perception might is essentially a top-down process (Bar, 2007; Hochstein and Ahissar, 2002; Nahum et al., 2008) with contextual information priming the detection of sensory patterns. Finally, it is also possible that listeners learned even the unfamiliar melodies through repetitions during the stimulus blocks (at least during those times when they heard the segregated organization). Either way, the fact that most listeners recognized the melodies afterwards (irrespective of whether or not they knew the songs from before the experiment) suggests that processes with access to long-term memory were engaged.

Effects of familiarity were shown by the significant interaction between the Group, Song and Melody factors. Hungarian participants were more likely to segregate sequences containing the Hungarian songs compared with English participants. Indeed, most of the Hungarian participants reported after the experiment that they knew the two Hungarian songs from before the experiment (Table 1). This effect cannot be explained by the acoustic cues, such as the smaller average pitch difference between consecutive notes, since the increase of segregation for Hungarian songs in the Hungarian listeners was observed in addition to the Melody main effect. Thus familiarity with a given sound pattern helps this pattern to be segregated from concurrent sounds. This conclusion is supported by the findings of Bey and McAdams (2002) that when a melody was introduced to listeners before the presentation of two interleaved sequences separately containing pitch patterns, participants were more likely to detect this melody and segregate the interleaved sequences (see also Devergie et al., 2010). Such schema-based influences on auditory streaming have been suggested in Bregman's (1990) framework as well as by Alain and Bernstein's (2008) more recent account.

The most dramatic effect found in the current study was that the presence of the original rhythm of the songs strongly promoted segregation, both by lengthening the *segregated* as well as by shortening the *integrated* perceptual phases. This result may indicate that rhythm acted both as a *percept-inducing* and as a *stabilizing cue* as it influenced both the competition between the percepts and the stability of the *segregated* percept (Denham and Winkler, 2006; Winkler et al., 2009a). A number of previous studies have shown that the auditory system detects rhythmic violations (see e.g., Geiser et al., 2009) even when the violations are not task-relevant (Ladinig et al., 2009). Further, neonates show brain responses suggesting that they are sensitive to gross rhythmic violations, such as the omission of the downbeat (Winkler et al., 2009b). Thus it is possible that rhythmic structure is extracted from a sound sequence even without voluntary effort, and that rhythm promotes the segregation of auditory streams. However the design of Experiment I included some confounding factors, allowing alternative interpretations of the results. First, the mean SOA was longer for the rhythmic than for the constant-SOA sequences – although this should promote integration, rather than segregation (van Noorden, 1975). Second, it is not clear whether participants actually processed the rhythmic structure of the sequences or whether the effect was due simply to SOA variation. Finally, it is also possible that segregation was caused by the overlap between tones from the two streams or by variation of the across-stream SOAs (shifting the tones from the halfway position between two successive tones in the other stream to random positions). Bregman and colleagues (Bregman et al., 2000) found that shorter temporal gaps and more overlap between high and low tones strongly promote segregation. In order to clarify which cue or cues were responsible for the effects observed in Experiment I, Experiment II was conducted. Because the significant interaction between the Rhythm and Song factors showed that the rhythm of the Hungarian songs had a more salient effect in both groups of listeners than the German songs, we only used the two Hungarian songs in Experiment II. In separate conditions, all of the above-mentioned confounds were tested: 1) mean SOA difference; 2) random SOA vs. rhythmic structure; 3) variation of the across-stream SOAs; and 4) overlap between tones belonging to different streams.

III. EXPERIMENT II

A. Methods

1. Participants

Participants were thirty-one young healthy Hungarian volunteers aged between 18 and 26 years (average age: 20.94 years, 17 female, 4 left handed). The experiment was conducted at the Institute of Cognitive Neuroscience and Psychology of the Research Centre for Natural Sciences, Hungarian Academy of Sciences. Inclusion criteria were identical to those reported for Experiment I.

2. Apparatus and stimuli

Similarly to Experiment I, four-minute sequences of the ABAB… structures were constructed from the two Hungarian songs H1 and H2. The parameters were as described for Experiment I, except for the following: The melody was randomized in each condition (Random-Melody conditions in terms of Experiment I). The SOA was either constant (conditions with constant rhythm, Figure 4, right panel) or variable (conditions with uneven rhythm, Figure 4 left panel).

In one of the six Constant Rhythm conditions, the SOA was 400 ms as it was in the Constant Rhythm conditions in Experiment I (Constant-Rhythm-400 condition). In the other five Constant-Rhythm conditions the SOA was 450 ms, setting it close to the combined mean SOA of the H1 and H2 songs (460.6 ms). In one of the latter Constant-Rhythm conditions the position of the B tone was set halfway between two A tones, as in all Constant-Rhythm conditions in Experiment I and the Constant-Rhythm-400 condition of Experiment II. This resulted in a uniform 225 ms across-stream SOA (Constant-Rhythm-450 condition). In two further Constant-Rhythm conditions, the B tones were shifted 20 ms forward or backward relative to the halfway position, resulting in 205 ms SOA between the A and B and 245 ms between B and A (Shifted-Forward condition), or 245 ms SOA between A and B and 205 ms SOA between B and A (Shifted-Backward condition). In the remaining two Constant-Rhythm conditions, the B tone was either shifted forward or backward by 112 ms resulting in either each B tone overlapping the preceding A tone by 87 ms (the A-B SOA being 113 ms and the B-A SOA 337 ms; Overlap-Forward condition) or the A tone overlapping the preceding B tone (the A-B SOA being 337 ms and the B-A SOA 113 ms; Overlap-Backward condition). In each case, the overlap was equal to the average overlap between the A and B tones in the Original Rhythm condition.

In the conditions with uneven rhythm (Figure 4, left panel), the order of the note durations either matched that of the original song (Original Rhythm) or was randomized (Mixed Rhythm). Finally, a condition with randomized SOA values (i.e., SOA could randomly take any value, not only those assigned to the note durations; Random Rhythm) was also administered. For the Random Rhythm sequences, SOA values were randomized while meeting the following criteria: a) the mean SOA values were 381.8 ms and 512.5 ms for H1 and H2, respectively, b) SOA values ranged from 200 to 800 ms, and c) the average overlap between the A and the B tones (87 ms) was identical to the Original Rhythm condition. This was achieved by starting from the two Original Rhythm stimulus sequences and time-shifting randomly selected tones by a random amount of time in a random direction (forward vs. backward), under three constraints. Firstly, tones were time-shifted in yoked pairs taken from the same song with the change in overlap caused by the shifting of one tone compensated by the corresponding shifting of the other tone. Secondly, only shifts resulting in an SOA within the SOA range of the Original Rhythm condition were allowed. Thirdly, only shifts keeping adjacent tones of the same sequence separated by at least 5 ms were allowed. Altogether, 600 pairs of shifts were applied to each of the two interleaved tone sequences. In order to prevent the survival of fragments of the original rhythm, priority was given to shifting tones that had not been shifted previously and when a tone was repeatedly shifted, the direction of the shift was restricted to that of the previous shift(s).

Streaming by melody and rhythm

The experimental session thus consisted of nine stimulus blocks (Original-Rhythm, Mixed-Rhythm, Random-Rhythm, Constant-Rhythm-400, Constant-Rhythm-450, Shifted-Forward, Shifted-Backward, Overlap-Forward, and Overlap-Backward) presented in a randomized order. The experimental procedures and data collection parameters were the same as reported for Experiment I.

3. Data analysis

The procedures for extracting perceptual phases and their parameters were identical to those described for Experiment I. Again, the overall proportions of the *both* and *neither* percepts were relatively small (both: 12.9%, neither: 1.0%); therefore, these were not analyzed further. The difference between the Constant-Rhythm-400 and the Constant-Rhythm-450 conditions was tested using Student's *t*-test. The effects of rhythm, overlap, and shifting was tested using a one-way repeated measures ANOVA of the Manipulation factor with eight levels corresponding to eight conditions (all, except for the Constant-Rhythm-400 condition), separately for the two main percepts (*integrated*, *segregated*) and for the proportions and log-mean phase durations. Sphericity violations were corrected using the Greenhouse Geisser correction of degrees of freedom. Post hoc tests were performed using Tukey HSD tests.

B. Results

1. Mean SOA difference

No significant differences were found between the Constant-Rhythm-400 and -450 conditions for either of the variables and percepts (p > .14).

Figure 5 shows the group-averaged proportion and log-mean phase duration values for the eight conditions included in the ANOVAs. The Manipulation factor had significant main effects in all four cases [F(7,210) = 16.78, ε = 0.498, $p < .001$, $\eta^2 = 0.359$ for the *integrated* proportions; F(7,210) = 14.87, ε = 0.562, $p < .001$, $\eta^2 = 0.331$ for the *segregated* proportion; F(7,210) = 5.446, ε = 0.629, $p < .001$, $\eta^2 = 0.154$ for the *integrated* log-mean phase durations; and F(7,210) = 3.128, ε = 0.680, $p < .05$, $\eta^2 = 0.094$ for the *segregated* log-mean phase durations].

2. Random SOA vs. rhythmic structure

For the *integrated* and *segregated* proportions and log-mean phase durations, post-hoc tests revealed that the three conditions where rhythm was not constant (original rhythm, random rhythm and mixed rhythm) did not differ from each other or from the overlap forward and backward conditions ($p > .12$).

3. Variation of the across-stream SOAs (shifting) and overlap between the tones

The two Shifted (Forward and Backward) conditions and the Constant-Rhythm-450 condition resulted in higher proportions of *integrated* percepts than the five other conditions (all $p < .01$), except that the Shifted-Forward condition was not significantly different from the Overlap-Forward condition ($p = .55$). Further, the Constant-Rhythm-450 condition and the two shifted conditions did not differ from each other (all $p > .29$). The same, but opposite direction effects were found for the *segregated* proportions; the only difference was that the Overlap-Backward condition was not significantly different from any of the other conditions ($p > .09$).

For *integrated* log-mean phase durations the Overlap-Forward condition did not differ from the three variable SOA conditions (all p > .97). In contrast, the three variable SOA conditions brought about significantly shorter *integrated* phases than either the Constant-Rhythm-450 or the Shifted-Forward condition (both p < .05). The Overlap-Backward and the Shifted-Forward conditions were not different from any of the other conditions (all p > .07). For *segregated* log-mean phase durations, only the Shifted-Forward condition resulted in significantly shorter phases than that obtained in the Mixed- and Random-Rhythm conditions (all p < .05).

C. Discussion of Experiment II

The results argue against some of the alternative explanations for the findings in Experiment I. The difference in the average SOA had no significant effect on the results. Also, no differences were obtained in our measures of the perceptual phases between the Original-Rhythm and the two randomized rhythm (Mixed- and Random-rhythm) conditions. This suggests that the rhythm effects found in Experiment I were probably not related to listeners using the musical rhythmic structure of the sequences as a cue of auditory stream segregation, because random (unstructured) but non-isochronous temporal structure had a similar effect. Further, the results obtained in these conditions didn’t significantly differ from those found for the Constant-Rhythm conditions with overlap between the tones from the two interleaved sequences. Because overlaps between tones belonging to the two interleaved sequences occurred in all variable-SOA conditions (Original, Mixed, and Random Rhythm conditions) these results indicate that in the various conditions with variable SOA, segregation was mainly promoted by tone overlap. This result is consistent with those of Bregman (2000).

On the surface, the current results contrast those of Devergie et al. (2010), Andreou et al. (2011), and Rimmele et al. (2012), who all found an increase of segregation when the to be ignored distractor sequence was delivered with a constant inter-onset interval (IOI) compared with when it was delivered with a random IOI. In these studies the degree of segregation was assessed in terms of how well listeners recognized a known melody (Devergie et al., 2010), detected an intensity pattern (Andreou et al., 2011), or detected an intensity-deviant sound embedded in one of the two interleaved tone sequences (Rimmele et al., 2012). Rimmele and colleagues (2012) found this only for young adults but not for elderly listeners. They also found a similar effect of regular IOI when the to-be-attended sequence was manipulated. In all three studies, the tones of the interleaved sequences either did not overlap each other in the compared conditions (Devergie et al., 2010 and Rimmele et al., 2012) or across-stream sound overlap was present in both compared conditions (Andreou et al., 2011). In the current study, the corresponding comparisons are those between the Overlap and the variable-SOA conditions (i.e., both conditions include across-stream overlap, as we have no variable-SOA condition without overlap). However, in contrast with Andreou et al. (2011), we found no significant increase of stream segregation for the Overlap compared with the variable-SOA conditions. The current paradigm differs from the three previous studies in two important ways. A) Participants in those studies were actively trying to segregate the interleaved sequences, because they were required to attend to one of them in order to perform their task. Thus participants may have used the strategy of actively suppressing the distractor sequence (which was helped by the isochronous compared with the jittered presentation of the distractors). In Rimmele et al.'s (2012) reversed condition (comparing between regular and random IOI in the to-be-attended sequence) listeners could have latched on to the regular timing of the sequence they followed. In contrast, in the current study, listeners had no task related specifically to either of the interleaved sequences and thus they may not have attempted to use all possible cues for segregating the two sequences. This difference in the paradigms therefore suggests that rhythmic regularities may only be used voluntarily as cues for stream segregation. B) The other difference between the current paradigm and those which showed that predictable timing helps stream segregation is that whereas we manipulated the temporal schedule of both interleaved sequences together (i.e., either both having variable or constant IOI), the three previous studies always delivered one sequence with random IOI and manipulated only the timing of the other sequence. Interleaving one sequence with regular IOI and another with random IOI could have biased the auditory system to utilize this cue for distinguishing the two sequences thus biasing the competition between the segregated and integrated sound organizations. This suggestion is compatible with that of Cusack et al. (2004), who proposed that qualitatively different streams are easier to segregate. The regular vs. irregular temporal schedule could have been acting as such a qualitative cue distinguishing the two interleaved sequences.

Shifting the tones (producing uneven across-stream SOAs) did not have a large effect on the perception of the sequences compared with the basic Constant-Rhythm condition. However, in some analyses the shifted sequences also did not differ from the overlap condition, suggesting that the results in these conditions varied more across listeners. It is possible that a larger shift, which still would not produce overlap between the interleaved tones, would have yielded perceptual patterns more similar to the overlap conditions.

One puzzling aspect of Experiment II is that the direction of the overlap appeared to have some effect on perception. When the ‘B’ tones commenced during the delivery of the ‘A’ tones (Overlap Forward) the results were similar to the three variable SOA conditions. However when ‘A’ tones commenced during the delivery of the ‘B’ tones (Overlap Backward) the effects of overlap were similar to those of the variable SOA only for the proportion of the *integrated* percept, but not for the other variables. In general, the ‘A’ and

‘B’ tones only differed from each other in their localization (ITD and IID). This may suggest that left-to-right and right-to-left overlaps are not processed fully symmetrically in the brain, perhaps because the right auditory cortex is more extensively involved in the processing of sound-source lateralization than the left one (Kaiser et al., 2000).

IV. GENERAL DISCUSSION AND CONCLUSIONS

The present experiments were designed to assess the effects of cues based on temporal structure on auditory stream segregation. We tested the effects of melody, rhythm, and familiarity by introducing tone patterns structured in various ways into two interleaved tone sequences. In order to avoid interactions between the effects of the various temporal and/or pitch-based regularities and the basic separation of the two interleaved sequences, the two interleaved sequences differed in ITD and IID (the main cues of sound lateralization) as opposed to frequency differences employed in most previous studies. This setup allowed the tunes to occupy largely overlapping pitch ranges, while producing balance between segregation and integration without the additional cues. A setup based on ITD and IID differences thus provides a good test-bed for future similar studies.

In Experiment I, we found that sequential pitch patterns structured in accordance with well-known rules (western musical conventions) as well as familiarity with the concrete patterns supports the segregation of auditory streams. These cues stabilize the *segregated* sound organization. Although, in contrast to Bendixen et al.’s earlier results (Bendixen et al., 2013; Bendixen et al., 2010) we found some signs that melodic pitch patterns also increase the likelihood for listeners to switch away from the *integrated* percept, the current results can also be explained by assuming a bias evoked by higher-level processes, such as checking for compliance with tonality.

Experiment II showed that the dramatic segregation-promoting effect of rhythmic structure found in Experiment I was either entirely or at least largely (with some possible effects of predictability, see Devergie et al., 2010) caused by overlaps between the tones of the two interleaved sequences. This result is compatible with those of Bregman et al. (2000) suggesting that overlap between two sounds acts as a strong cue of auditory stream segregation, both helping to reject the *integrated* sound organization and stabilizing the *segregated* one.

In summary, the presence of both melody and rhythm, separately inserted into two interleaved tone sequences promoted segregation of two sequences. In general, both of these temporal-structure-based cues increased the proportion of segregation and they did so by extending the duration of the intervals during which listeners heard two streams as well as shortening the intervals during which they heard a single stream. Thus these cues can both induce and stabilize auditory streams. Both the effects of melody and rhythm have been further qualified. The effects of melody are stronger for familiar tunes suggesting the involvement of long-term memory in auditory scene analysis. The effects of rhythm have been found to be mediated by overlap between the tones of the two streams. However, two issues remain open. Firstly, it is possible that well-learned rules, such as tonality may act as *percept-inducing cues*, perhaps by extending what is regarded as “similar” by the brain. Secondly, it is possible that some temporal-structure-based cues bias auditory scene analysis by engaging processes operating outside the primary mechanisms of auditory stream segregation (cf., Alain and Bernstein, 2008).

ACKNOWLEDGEMENTS

This study was funded by the EU FP7-ICT-231168-SCANDLE project (acoustic SCene ANalysis for Detecting Living Entities), by the Lendület project awarded to István Winkler by the Hungarian Academy of Sciences (contract number LP2012-36/2012), by the German Research Foundation (Cluster of Excellence DFG 1077 "Hearing4all"), by the German Academic Exchange Service (Deutscher Akademischer Austauschdienst [DAAD], Project 56265741), and by the Hungarian Scholarship Board (Magyar Ösztöndíj Bizottság [MÖB], Project 39589).

Alain, C., and Bernstein, L. J. (**2008**). "From sounds to meaning: The role of attention during auditory scene analysis," Curr. Opin. Otolaryngol. Head. Neck. Surg. **16**, 485–489.

Andreou, L. V., Kashino, M., and Chait, M. (**2011**). "The role of temporal regularity in auditory segregation," Hear. Res. **280**, 228-235.

Bar, M. (**2007**). "The proactive brain: Using analogies and associations to generate predictions," Trends Cogn. Sci. **11**, 280-289.

Bendixen, A., Bőhm, T., Szalárdy, O., Mill, R., Denham, S. L., and Winkler, I. (**2013**). "Different roles of similarity and predictability in auditory stream segregation," Learn. Percept. **5**, 37-54.

Bendixen, A., Denham, S. L., Gyimesi, K., and Winkler, I. (**2010**). "Regular patterns stabilize auditory streams," J. Acoust. Soc. Am. **128**, 3658-3666.

Bey, C., and McAdams, S. (**2002**). "Schema-based processing in auditory scene analysis," Percept. Psychophys. **64**, 844-854.

Bőhm, T. M., Shestopalova, L., Bendixen, A., Andreou, A. G., Georgiou, J., Garreau, G., Poliquen, P., Cassidy, A., Denham, S. L., and Winkler, I. (**2013**). "The role of perceived source location in auditory stream segregation: Separation affects sound organization, common fate does not," Learn. Percept. **5**, 55-72.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA), pp. 47-184, 411-453.

Bregman, A. S., Ahad, P. A., Crum, P. A., and O'Reilly, J. (**2000**). "Effects of time intervals and tone durations on auditory stream segregation," Atten. Percept. Psychophys. **62**, 626-636.

Cuddy, L. L. (**1991**). "Melodic patterns and tonal structure: Converging evidence," Psychomusic. **10**, 107-126.

Cusack R., Deeks J., Aikman G., Carlyon R. P. (**2004)**. "Effects of location, frequency region, and time course of selective attention on auditory scene analysis," J. Exp. Psychol. Hum. Percept. Perform. **30**, 643-656.

Demany, L., and Semal, C. (**2002**). "Limits of rhythm perception," Q. J. Exp. Psychol. A. **55**, 643-657.

Denham, S. L., Gyimesi, K., Stefanics, G., and Winkler, I. (**2010**). "Stability of perceptual organisation in auditory streaming," in *The Neurophysiological Bases of Auditory Perception.*, edited by E. A. Lopez-Poveda, A. R. Palmer & R. Meddis (Springer, New York), pp. 477-487.

Denham, S. L., Gyimesi, K., Stefanics, G., and Winkler, I. (**2013**). "Perceptual bistability in auditory streaming: How much do stimulus features matter?," Learn. Percept. **5**, 73-100.

Denham, S. L., and Winkler, I. (**2006**). "The role of predictive models in the formation of auditory streams," J. Physiol. Paris **100**, 154-170.

Devergie, A., Grimault, N., Tillmann, B., and Berthommier, F. (**2010**). "Effect of rhythmic attention on the segregation of interleaved melodies," J. Acoust. Soc. Am. **128**, EL1-EL7.

Dowling, W. J. (**1973**). "The perception of interleaved melodies," Cog. Psychol. **5**, 322-337.

Dowling, W. J., Lung, K. M., and Herrbold, S. (**1987**). "Aiming attention in pitch and time in the perception of interleaved melodies," Percept. Psychophys. **41**, 642-656.

French-St George, M., and Bregman, A. S. (**1989**). "Role of predictability of sequence in auditory stream segregation," Percept. Psychophys. **46**, 384-386.

Geiser, E., Ziegler, E., Jancke, L., and Meyer, M. (**2009**). "Early electrophysiological correlates of meter and rhythm processing in music perception," Cortex **45**, 93-102.

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., and Oxenham, A. J. (**2005**). "Neuromagnetic correlates of streaming in human auditory cortex," J. Neurosci. **25**, 5382-5388.

Hochstein, S., and Ahissar, M. (**2002**). "View from the top: hierarchies and reverse hierarchies in the visual system," Neuron **36**, 791-804.

Jones, M. R. (**1976**). "Time, our lost dimension: toward a new theory of perception, attention, and memory," Psychol. Rev. **83**, 323-355.

Jones, M. R., and Boltz, M. (**1989**). "Dynamic attending and responses to time," Psychol. Rev. **96**, 459-491.

Kaiser, J., Lutzenberger, W., Preissl, H., Ackermann, H., and Birbaumer, N. (**2000**). "Right-hemisphere dominance for the processing of sound-source lateralization," J. Neurosci. **20**, 6631-6639.

Koelsch, S., and Siebel, W. A. (**2005**). "Towards a neural basis of music perception," Trends Cogn. Sci. **9**, 578-584.

Ladinig, O., Honing, H., Háden, G. P., and Winkler, I. (**2009**). "Probing attentive and pre-attentive emergent meter in adult listeners without extensive music training," Music Percept. **26**, 377-386.

Leopold, D. A., and Logothetis, N. K. (**1999**). "Multistable phenomena: Changing views in perception," Trends Cogn. Sci. **3**, 254-264.

McDermott, J. H.,Wrobleski, D., and Oxenham, A. J. (**2011**). "Recovering sound sources from embedded repetition, " Proc. Natl. Acad. Sci. USA **108**, 1188–1193.

Mill, R. W., Bőhm, T. M., Bendixen, A., Winkler, I., and Denham, S. L. (**2013**). "Modelling the emergence and dynamics of perceptual organisation in auditory streaming," PLoS Comp. Biol. **9**, e1002925.

Moore, B. C. J., and Gockel, H. (**2002**). "Factors influencing sequential stream segregation," Acta Acust. United. Ac. **88**, 320-333.

Moreno-Bote, R., Shpiro, A., Rinzel, J., and Rubin, N. (**2010**). "Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance," J. Vision **10**, 1-18.

Nahum, M., Nelken, I., and Ahissar, M. (**2008**). "Low-level information and high-level perception: the case of speech in noise," PLoS Biol. **6**, e126.

Pressnitzer, D., and Hupé, J-M. (**2006**). "Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization," Curr. Biol. **16**, 1351-1357.

Rimmele, J., Schröger, E., and Bendixen, A. (**2012**). "Age-related changes in the use of regular patterns for auditory scene analysis," Hear. Res. **289**, 98-107.

Rogers, W. L., and Bregman, A. S. (**1993**). "An experimental evaluation of three theories of auditory stream segregation," Percept. Psychophys. **53**, 179-189.

Snyder, J. S., and Alain, C. (**2007**). "Toward a neurophysiological theory of auditory stream segregation," Psychol. Bull. **133**, 780-799.

Snyder, J. S., Gregg, M. K., Weintraub, D. M., and Alain, C. (**2012**). "Attention, awareness, and the perception of auditory scenes," Front. Psychol. **3**, 15.

Snyder, J. S., and Weintraub, D. M. (**2011**). "Pattern specificity in the effect of prior Δf on auditory stream segregation.," J. Exp. Psychol. Hum. Percept. Perform. **37**, 1649-1656.

Sussman, E., Ritter, W., and Vaughan, H. G., Jr. (**1998**). "Attention affects the organization of auditory input associated with the mismatch negativity system," Brain Res. **789**, 130-138.

Szalárdy, O., Bendixen, A., Tóth, D., Denham, S. L., and Winkler, I. (**2013**). "Modulation-frequency acts as a primary cue for auditory stream segregation," Learn. Percept. **5**, 149-161.

Trainor, L. J., and Trehub, S. E. (**1994**). "Key membership and implied harmony in Western tonal music: developmental perspectives," Atten. Percept. Psychophys. **56**, 125-132.

van Noorden, L. P. A. S. (**1975**)."Temporal coherence in the perception of tone sequences," Doctoral dissertation, Technical University Eindhoven, Eindhoven

Winkler, I., Denham, S., Mill, R., Bőhm, T. M., and Bendixen, A. (**2012**). "Multistability in auditory stream segregation: a predictive coding view," Philos. Trans. R. Soc. Lond. B Biol. Sci. **367**, 1001-1012.

Winkler, I., Denham, S. L., and Nelken, I. (**2009a**). "Modeling the auditory scene: predictive regularity representations and perceptual objects," Trends Cogn. Sci. **13**, 532-540.

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., and Honing, H. (**2009b**). "Newborn infants detect the beat in music," Proc. Natl. Acad. Sci. USA **106**, 2468-2471.

TABLE I. The number and percentage of participants who were familiar with and/or recognized one or both Hungarian and/or German songs in the Hungarian (top row) and in the English (bottom row) group.

	Familiarity		**Recognition**	
	Hungarian songs	German songs	Hungarian songs	German songs
Hungarian	22(100%)	3(13.5%)	18(81.8%)	13(59.1%)
English	0	0	18(85.7%)	14(66.7%)

FIGURE 1. Musical scores for each of the four songs used in Experiment I.

FIGURE 2. Percept proportions. Each panel shows percept proportions obtained for the Hungarian (left column within each panel) and the English group (right column within each panel), respectively. Except for the bottom right panel, rows within each panel show the three possible Song combinations (Hungarian-Hungarian, German-German, Hungarian-German. The top-left panel compares the Constant-Rhythm/Random-Melody (left) and Constant-Rhythm/Original-Melody (right) conditions, thus illustrating the effects of the Melody factor. The bottom-left panel compares the Constant-Rhythm/Random-Melody (left) and Original-Rhythm/Random-Melody (right) conditions, thus illustrating the effects of the Rhythm factor. The top-right panel shows the effects of Rhythm and Melody together by comparing between the Constant-Rhythm/Random-Melody (left) and Original-Rhythm/Original-Melody (right) conditions. The bottom-left panel shows the same three comparisons but with the Song factor collapsed: Melody (upper row), Rhythm (middle row), and Melody and Rhythm together (bottom row).

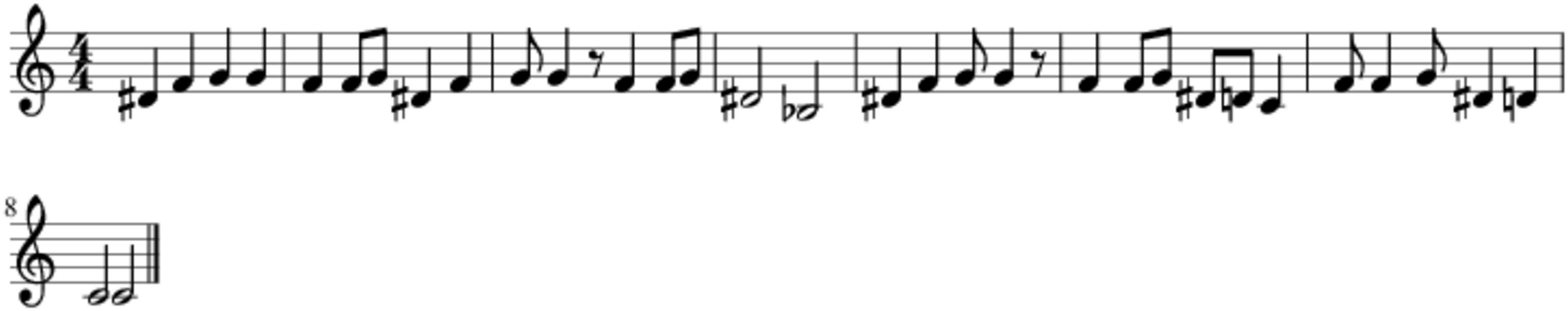
FIGURE 3. Log-mean perceptual phase durations. Each panel shows log-mean perceptual phase durations for the Hungarian (left column within each panel) and the English group (right column within each panel), respectively. Except for the bottom right panel, rows within each panel show the three possible Song combinations (Hungarian-Hungarian, German-German, Hungarian-German. The top-left panel compares the Constant-Rhythm/Random-Melody (left) and Constant-Rhythm/Original-Melody (right) conditions, thus illustrating the effects of the Melody factor. The bottom-left panel compares the Constant-Rhythm/Random-Melody (left) and Original-Rhythm/Random-Melody (right) conditions, thus illustrating the effects of the Rhythm factor. The top-right panel shows the effects of Rhythm and Melody together by comparing between the Constant-Rhythm/Random-Melody (left) and Original-Rhythm/Original-Melody (right) conditions. The bottom-left panel shows the same three

comparisons but with the Song factor collapsed: Melody (upper row), Rhythm (middle row), and Melody and Rhythm together (bottom row).

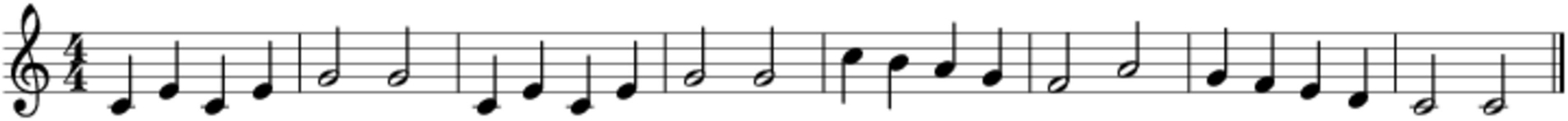
FIGURE 4. Schematic illustration of the experimental conditions. The left panels illustrate the non-isochronous experimental conditions. The Original Rhythm and Mixed Rhythm conditions contain the same silent intervals but their order is different. In the Random Rhythm condition, the average overlap between the left and right tones was set to match the Original Rhythm condition, whereas the duration of the silent intervals varied continuously between the shortest and longest interval of the Original Rhythm condition. The right panels show the isochronous conditions.

FIGURE 5. Percept proportions and log-mean perceptual phase durations. The top panel shows the proportions while the bottom panel shows the log-mean phase durations. The rows show the two kinds of manipulations: Rhythm (Constant-Rhythm-450, Original-Rhythm, Mixed-Rhythm, and Random-rhythm) and Shift/Overlap (Shifted-Forward, Shifted-Backward, Overlap-Forward, and Overlap-Backward).

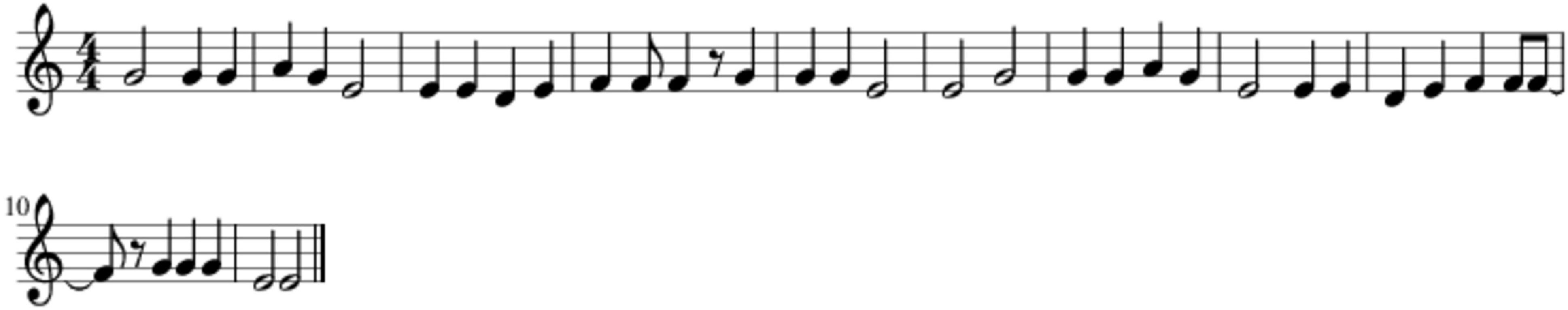
H1



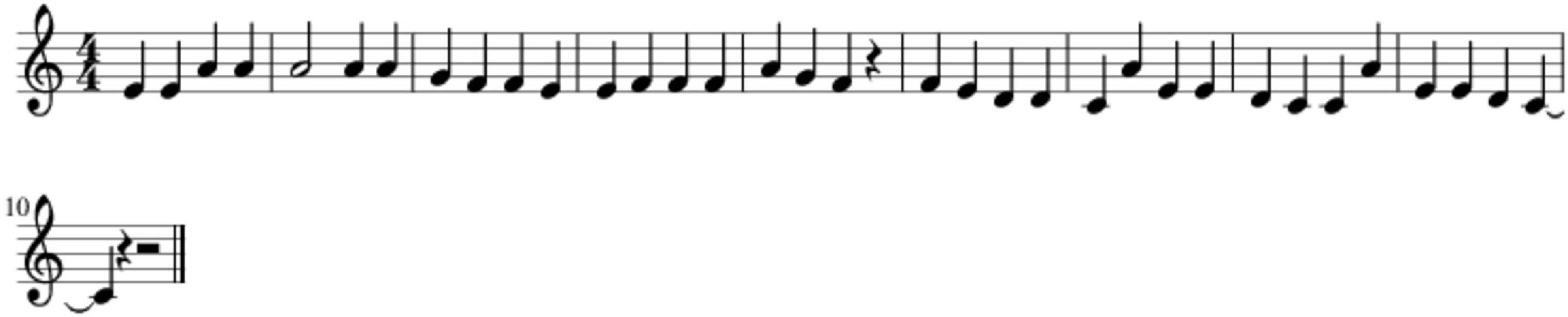
H2

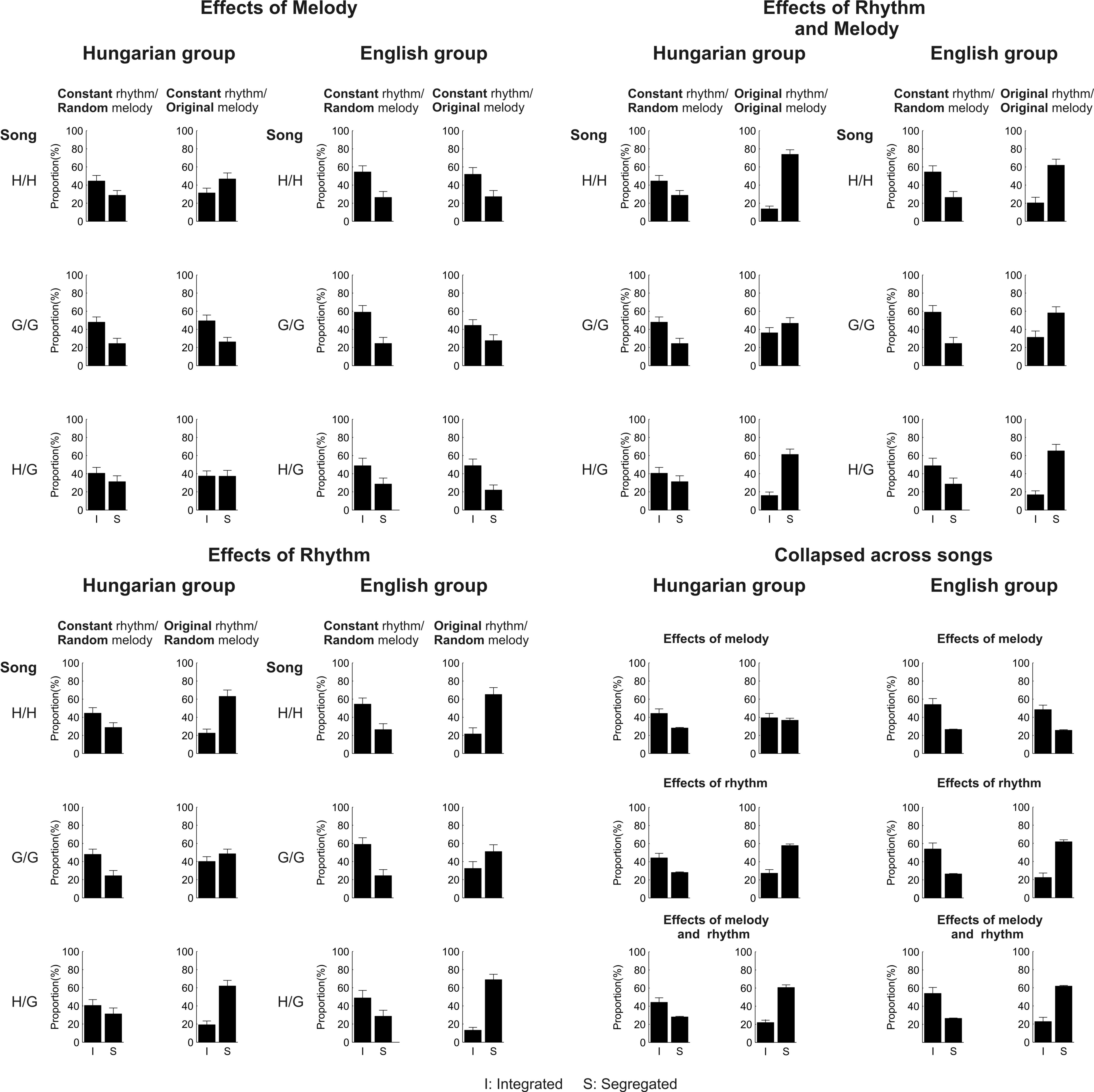


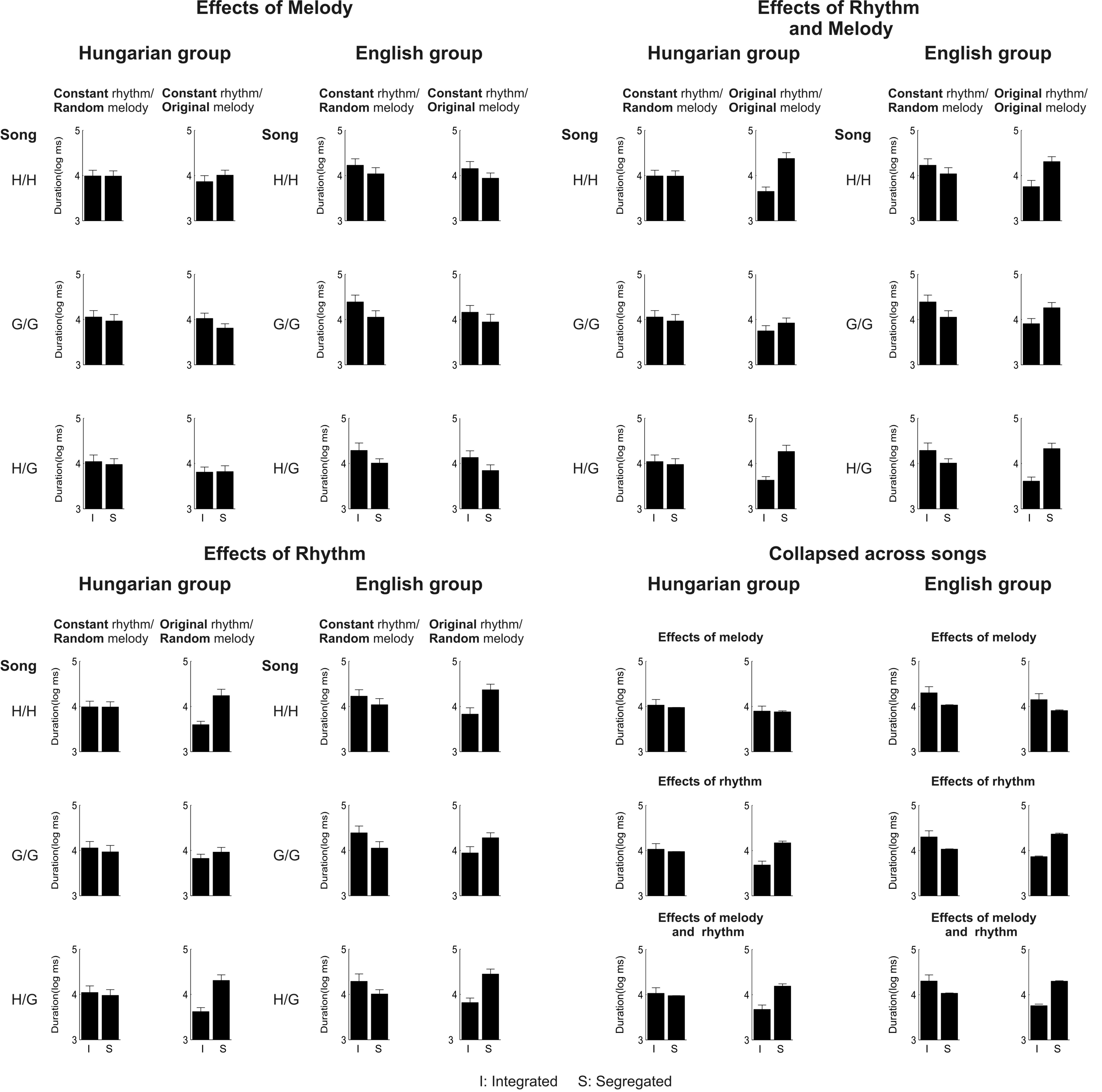
G1

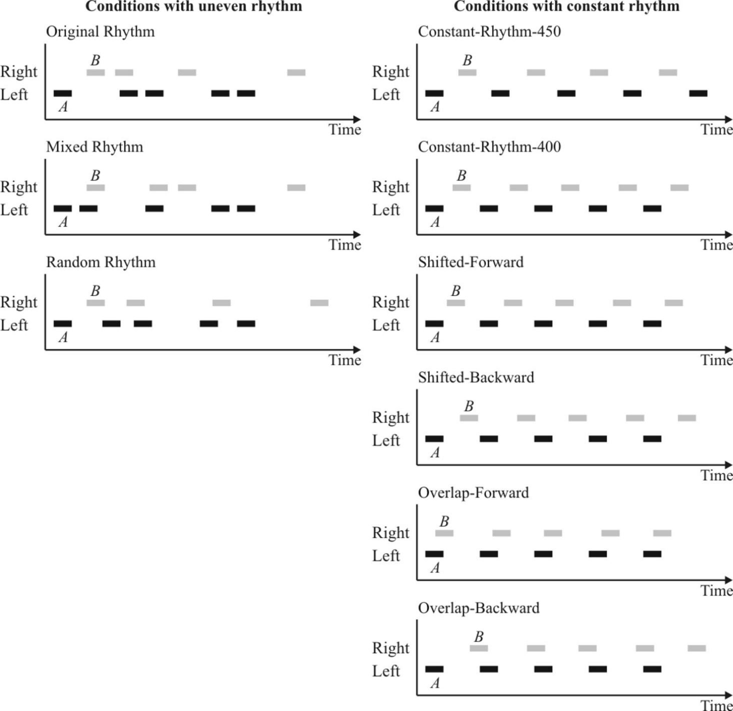


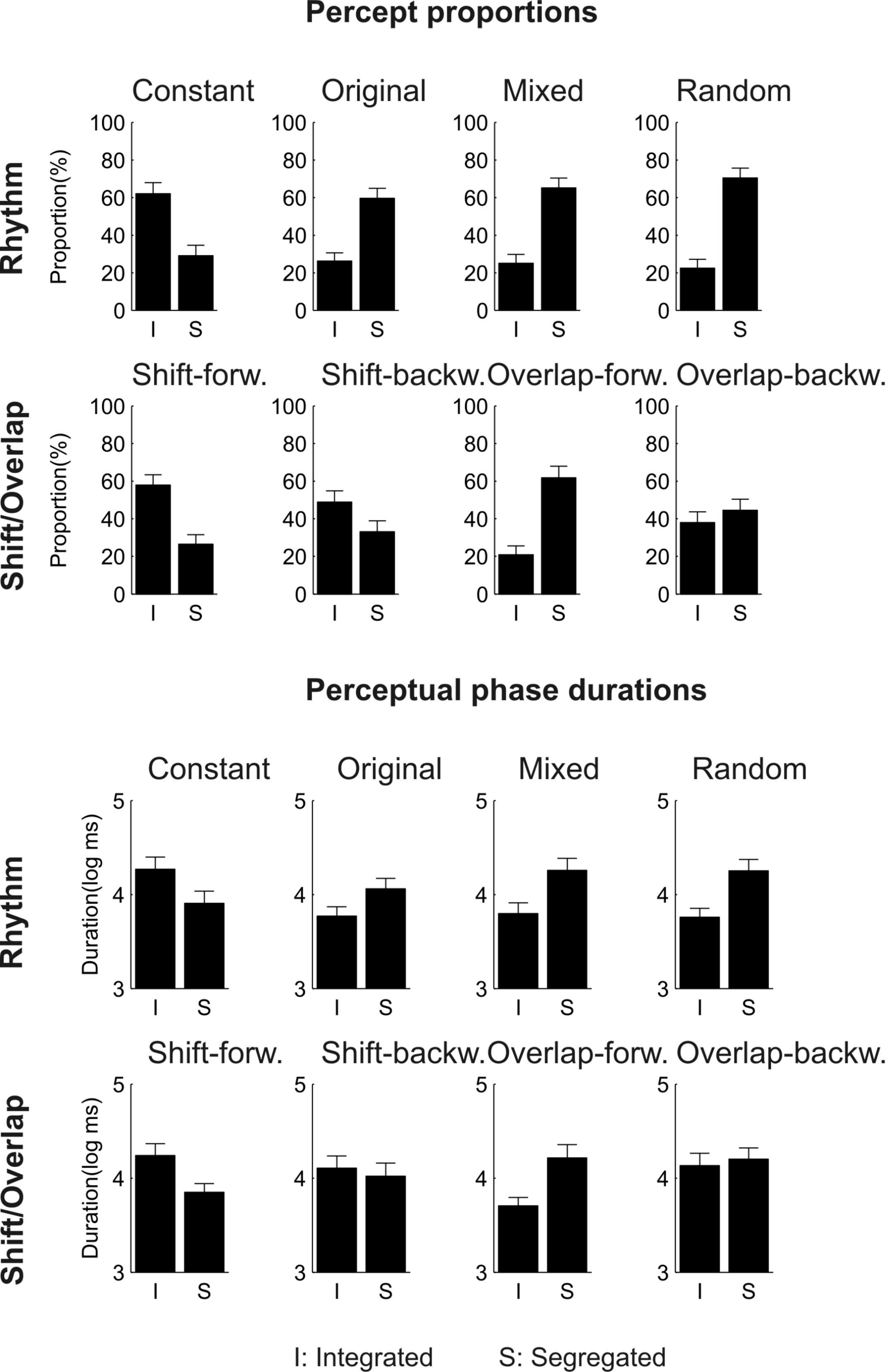
G2




Effects of Melody
Hungarian group
English group
Constant rhythm/ Random melody
Constant rhythm/ Original melody
Effects of Rhythm and Melody
Hungarian group
English group
Constant rhythm/ Random melody
Original rhythm/ Original melody
Song
H/H
G/G
H/G
Proportion(%)
0
20
40
60
80
100
I
S
Effects of Rhythm
Hungarian group
English group
Constant rhythm/ Random melody
Original rhythm/ Random melody
Collapsed across songs
Hungarian group
English group
Effects of melody
Effects of rhythm
Effects of melody and rhythm
I: Integrated
S: Segregated


Effects of Melody
Effects of Rhythm and Melody
Effects of Rhythm
Collapsed across songs
Hungarian group
English group
Constant rhythm/ Random melody
Constant rhythm/ Original melody
Original rhythm/ Original melody
Original rhythm/ Random melody
Song
H/H
G/G
H/G
Duration(log ms)
3
4
5
I
S
Effects of melody
Effects of rhythm
Effects of melody and rhythm
I: Integrated
S: Segregated


Conditions with uneven rhythm
Original Rhythm
B
Right
Left
A
Time
Mixed Rhythm
B
Right
Left
A
Time
Random Rhythm
B
Right
Left
A
Time
Conditions with constant rhythm
Constant-Rhythm-450
B
Right
Left
A
Time
Constant-Rhythm-400
B
Right
Left
A
Time
Shifted-Forward
B
Right
Left
A
Time
Shifted-Backward
B
Right
Left
A
Time
Overlap-Forward
B
Right
Left
A
Time
Overlap-Backward
B
Right
Left
A
Time


Percept proportions
Constant
Original
Mixed
Random
Rhythm
Proportion(%)
100
80
60
40
20
0
I
S
Shift-forw.
Shift-backw.
Overlap-forw.
Overlap-backw.
Shift/Overlap
Perceptual phase durations
Duration(log ms)
5
4
3
I: Integrated
S: Segregated