Different Voice Part Perceptions in Polyphonic and Homophonic Musical Textures

Short title: VOICE PART PERCEPTION IN MUSIC

Kai Ishida[1, 2], Hiroshi Nittono[1]

[1] Graduate School of Human Sciences, Osaka University, Osaka, Japan

[2] Japan Society for the Promotion of Science, Tokyo, Japan

Corresponding author: Kai Ishida, Graduate School of Human Sciences, Osaka University, 1-2 Yamadaoka, Suita, Osaka 565-0871, JAPAN; ishida@hus.osaka-u.ac.jp; ORCID: 0000-0001-6485-0950

Hiroshi Nittono, Graduate School of Human Sciences, Osaka University, 1-2 Yamadaoka, Suita, Osaka 565-0871, JAPAN; nittono@hus.osaka-u.ac.jp; ORCID: 0000-0001-5671-609X

Authorship contribution statement: Kai Ishida: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing - original draft Project administration, Funding acquisition. **Hiroshi Nittono:** Conceptualization, Methodology, Formal analysis, Funding acquisition, Writing - review & editing.

Conflict of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding information: This word was supported by JSPS KAKENHI Grant Number JP22J20773.

Keywords: redundant signals effect (RSE), race model inequality (RMI), harmony, high-voice superiority effect (HVSE), polyphony

Word count: 6290

Data accessibility:

The sound materials and datasets used and analyzed for the present paper can be made available at https://osf.io/agux4/.

Abstract

Separate voice part perception has been shown in polyphonic music. However, it remains unclear whether this segregation of voice parts is specific to polyphony or also occurs in homophonic music. The present study compared voice part perceptions in polyphony and homophony using a redundant signals effect (RSE) paradigm. The RSE means that reaction times are shorter for two simultaneously presented signals than for one of these signals. At the final position of the four-voice homophonic and polyphonic sequences, notes in two voice parts were altered to out-of-key notes independently or simultaneously. Participants ($N = 208$) responded to any deviant tones while withholding responses to non-deviant tones. All combinations of deviant voice parts (i.e., soprano–bass, tenor–bass, and alto–tenor) elicited RSEs in polyphonic and homophonic sequences, suggesting separate voice part perception, irrespective of musical texture. However, evidence of the coactivation of separate perceptual modules was obtained only for polyphonic sequences. Deviants in higher voice parts were detected faster and more accurately than those in lower voice parts in both musical textures. These results indicate that voice parts are perceived separately, with a bias toward higher voice parts in both musical textures, but voice parts are more segregated in polyphony.

In Western tonal music, tunes are composed of various types of musical textures, such as polyphony, homophony, and heterophony. Of these, voice perception in polyphony has been the most intensively examined. For instance, score analyses have revealed composers' endeavors to enhance the independence of voice parts in polyphony (Huron, 1991, 2008; Huron & Fatini, 1989). Behavioral (Dowling, 1973; Gregory, 1990; Saup et al., 2010; Sloboda & Edowrthy, 1981) and neurophysiological (Fujioka et al., 2005; Hubeth & Fujioka, 2017; Janata et al., 2002; Marie et al., 2012) studies have also demonstrated separate perceptions of each voice part in polyphony. The separate voice part perception is an example of the auditory stream segregation, a perceptual organization process that groups sound input into meaningful streams based on auditory cues such as frequency (Bregman & Campbell, 1971; Bregman et al., 1990) and timing (Vos, 1995).

The perception of voice parts in polyphony has often been examined in the framework of attentional theories (Barrett et al., 2021; Bigand et al., 2000; Disbergen et al., 2018; Gregory, 1990; Hausfeld et al., 2021; Saup et al., 2010). For example, Gregory (1990) proposed a divided attention account, which suggests that melody lines in polyphony can be simultaneously perceived in parallel, by showing that participants recognized separate single melody lines of polyphonic music above the chance level. Another perspective is the figure–ground account, which suggests that listeners focus on one melody of polyphonic music as the figure while staying aware of the other melodies in the background. This theory was proposed by Sloboda and Edworthy (1981), who showed that error detection performance in well-learned polyphonic music pieces was higher when the melodies were in the same key than when the melodies were in different keys. This result can be explained by considering that the figure–ground

representation was constructed in terms of harmony when the melodies were in the same key, but not when they were in different keys. Using a similar paradigm, Bigand et al. (2000) proposed the integrative model, which suggests that listeners tend to integrate voice parts into a single perceptual object to compensate for attentional limitations. These attentional accounts argue that multi-voice parts in polyphony are perceived separately, but to some extent, they are processed integratively.

Neurophysiological studies using the event-related potential (ERP) have also demonstrated the separate processing of voice parts in polyphony. Fujioka et al. (2005) presented two melodies simultaneously, each with an infrequent pitch deviant to which a mismatch negativity (MMN) was recorded. The MMN is an ERP component that reflects auditory deviance detection based on sensory memory traces of auditory events (Näätänen et al., 2005). Deviants in each voice part elicited an MMN even when the simultaneously presented tones were consonant as a whole, suggesting separate voice processing for each voice part based on separate sensory memory traces. This segregation has been observed even in non-musicians and when the stimuli are ignored (Fujioka et al., 2005; Marie et al., 2012).

Furthermore, although several voice parts are perceived separately, higher and lower voices may have different superiorities. For example, the higher voice part is perceived more prominently than the lower voice part with respect to pitch information (Fujioka et al., 2005; Huberth & Fujioka, 2017; Marie et al., 2012), while the lower voice part is superior for encoding temporal information (Hove et al., 2014). The biased effect of the higher voice is referred to as the *high-voice superiority effect* (HVSE), and previous ERP studies have reported enhanced MMN amplitude in the pitch deviant in the higher voice part compared to the lower voice part (Fujioka et al., 2005; Marie et al.,

2012). Therefore, neural evidence for separate voice part perception in the pitch dimension has been demonstrated with a bias toward higher voice parts. As discussed in Trainer et al. (2014), saliency in higher pitch perception may be due to different activation patterns for different frequencies in the peripheral auditory system, such as the middle ear and cochlea (Békésy, 1960).

However, there is an unresolved issue. While previous studies have proposed the separate processing of polyphonic melodies in an integrative manner (Barretto et al., 2018; Bigand et al., 2000; Sloboda & Edworthy, 1981), it is unclear whether this nature of voice part perception is specific to polyphony. For example, the voice leading in the harmonic chord progression of homophony can serve as a cue for stream segregation, because it creates continuity and makes the streams of individual voice lines distinct based on perceptual principles (Huron, 2001). To answer this question, a comparison between different music styles is required. In homophony, each voice part may be perceived in a more fused manner compared to polyphony, because all voice parts construct a unit of harmony moment by moment. Consequently, separate voice perception occurs in polyphony but not in homophony. The present study aimed to examine whether separate voice part perception is specific to polyphony.

To address this question, this study introduced a new experimental protocol, the redundant signals effect (RSE) paradigm, which is based on reaction times (RTs) in a simple target detection task using multidimensional signals. Because previous behavioral studies have primarily used memory tasks (Bigand et al., 2000; Gregory, 1990; Sloboda & Edworthy, 1981), the results may have depended on cognitive abilities in encoding and retrieval, which are not directly related to music perception. To eliminate these effects, the present study focuses on the RSE, which means that RTs to

target signals are shorter when two different signals are presented simultaneously than when only one of the signals is presented. The RSE has been accounted for by a race model and a coactivation model (Miller, 1982). The race model suggests statistical facilitation, in which responses in the redundant signals condition are caused by the fastest single signal processing (Raab, 1962). The coactivation model suggests that activations from different channels are combined to initiate a faster response (Miller, 1982, 1986, 2004; Miller & Ulrich, 2003). In the race model, the predicted redundancy gain follows the race model inequality (RMI) when assuming that signals are auditory and visual stimuli (Miller, 1982):

$$F_{AV}(t) \leq F_A(t) + F_V(t)$$

for every value of t, in which F_A and F_V are the cumulative distribution functions (CDFs) of the RTs in the auditory and visual signal conditions, and F_{AV} is the CDF of the RT in the redundant signals condition. When the RMI is not violated, the race model is upheld, and when the RMI is violated, the race model is rejected, and a coactivation model is adopted, assuming context invariance (Gondan & Minakata, 2016; Luce, 1986).

Preventing the fusion of two signals into a single percept is a prerequisite for the occurrence of the RSE (Schröter et al., 2007). Thus, the presence of the RSE could be an indicator of separate voice part processing, while the absence of the RSE could be an indicator of fused voice part processing. Moreover, evidence of coactivation is observed when two sufficiently distinct signals are manipulated in one perceptual object (Fiedler et al., 2011; Mordkoff & Danek, 2011; Mordkoff & Yantis, 1993; Schröter et al., 2007). For example, in the visual domain, redundant signals consisting of signals in shape (e.g., X) and color (e.g., green) elicited coactivation, while redundant signals

consisting of signals in two colors (e.g., green and red) elicited statistical facilitation (Mordkoff & Yantis, 1993). As examples in the auditory domain, Fiedler et al. (2011) reported coactivation elicited by the signals in frequency and location dimensions, while Schröter et al. (2007) failed to observe an RSE when the signals were two pure tones with different frequencies. Therefore, evidence of coactivation could indicate that voice parts are separate enough to be processed in separate perceptual modules (Fiedler et al., 2011; Mordkoff & Danek, 2011; Mordkoff & Yantis, 1993).

In the present study, voice part perception was compared between four-voice polyphonic and homophonic sequences using the RSE. In the final chord of each sequence, one or both of any two of the four voice parts were occasionally deviated to an out-of-key note (i.e., pitch deviant). Three combinations (i.e., conditions) of target voice parts were examined: soprano and bass as a combination of outer voices, tenor and bass as a combination of outer and inner voices, and alto and tenor as a combination of inner voice parts. Participants were asked to detect all deviant chords while withholding their responses to non-deviant chords (i.e., the Go/NoGo task). If the separate voice part perception is specific to polyphony, the RSE as evidence of separate signal processing should be observed by all deviant combinations in the polyphonic sequence but not in the homophonic sequence. Additionally, the present study compared detection performance between higher and lower voice parts to examine the HVSE. If the higher voice part is more prominently perceived in music processing, the mean RT in the higher voice part will be shorter than in the lower voice part, and the hit rate will be higher in the higher voice part than in the lower voice part.

Methods

All deviant conditions were preregistered as separate experiments before sampling. The preregistration details for each experiment can be found at the following links: the soprano–bass deviant condition (Experiment 1, https://osf.io/5jdrc), the tenor–bass deviant condition (Experiment 2, https://osf.io/s7dhk), and the alto–tenor deviant condition (Experiment 3, https://osf.io/p8v7d). The protocols were approved by the Behavioral Research Ethics Committee of the Osaka University School of Human Sciences, Japan (HB022-062 for the soprano–bass deviant condition and HB022-101 for the tenor–bass and alto–tenor deviant conditions), and informed consent was obtained from all participants.

Sample Size Calculation

In our previous study using similar musical sequences (Ishida & Nittono, 2023), the RSE had an effect size of Cohen's $d = 0.832$. A power analysis using G*power (Faul et al., 2007) resulted in $N = 14$ for a paired t-test ($\alpha = .05$ and $1-\beta = .90$, one-sided). We also conducted a sample size analysis for the RMI test based on our previous study's data to ensure sufficient power for comparing the double-deviant CDF and each summed single-deviant CDF. A significant violation of the RMI was observed at the first five decile points when the data of our previous study were analyzed based on Miller's method, which compared the mean reaction times (MRTs) corresponding to each decile of CDFs (summed CDFs vs. double-deviant CDFs) using a paired t-test. Among the first five decile points, the smallest effect size was found at the 5th decile point, $dz = -0.476$. The required sample size was obtained from the power contour (Baker et al., 2021), which is a function of the number of trials and the sample size, given a mean difference, between-participant standard deviation, and within-participant standard deviation. In the current case, a mean difference of −11.57 ms (i.e., MRT

difference between the summed and double-deviant CDFs at the 5th decile points), a between-participants standard deviation of 24.31 ms at the 5th decile point, and a within-participant standard deviation of 50, which is considered sufficiently large, were applied. The results showed that we required $N \geq 58$ to obtain power $1-\beta > .90$ for 24 trials (the minimal number of trials to be included in the analysis). A post-hoc simulation also confirmed the validity of this sample size (see Supplementary Material).

We also conducted a sample size analysis for the HVSE test. A power analysis using G*power resulted in $N = 36$ for a paired t-test (Cohen's $d = 0.5$, $\alpha = .05$ and $1-\beta = .90$, one-sided). A sample size of 58 or more was determined by taking the larger sample size in these calculations. For the soprano–bass deviant condition, considering the possibility of outliers and missing values, 90 participants were recruited. However, for the tenor–bass and alto–tenor deviant conditions, data were collected from 116 participants, which is 200% of the minimum sample size. This larger sample size was chosen due to the expected higher dropout rate because of the difficulty of deviance detection for lower voices. The participants received 880 Japanese yen in all deviant conditions as an honorarium. None of the participants reported hearing impairments.

Participants

The detailed attributes of the participants are described in Supplementary Material. For the analyses of the voice part perception, participants with a hit rate lower than 80% for at least one deviant voice and those with mismatches in gender and age data before and after the experiment were excluded. Consequently, the data of 64, 76, and 68 participants were used for hypothesis testing in the soprano–bass, tenor–bass, and alto–tenor deviant conditions, respectively.

In the analysis of the HVSE, participants with a hit rate lower than 50% for at

least one deviant voice and those with mismatches in gender and age data before and after the experiment were excluded. Because the HVSE analysis was based on mean RTs, which should be more stable than single-trial RTs for the RSE analysis. The 50% criterion was set because it was the chance level. Note that virtually the same results were obtained when only the participants included in the RSE analysis were analyzed (see Supplementary Material). Consequently, 82, 99, and 111 participants were used for hypothesis testing in the soprano–bass, tenor–bass, and alto–tenor deviant conditions, respectively.

Figure 1

Stimuli

Examples of the homophonic and polyphonic sequences are depicted in Figure 1. For the examination of the soprano–bass and tenor–bass conditions, three types of homophonic sequences with different chord inversions and three types of polyphonic sequences with four independent melodic parts were composed, such that both sequences adhered to Western harmony rules (I→IV→II→V→I). These sequences were played with a piano timbre. The homophonic sequence consisted of five half-note chords, while the polyphonic sequence consisted of different notes, from half notes to eighth notes. However, the final chord of both sequences was 1,200 ms, and the overall duration of each sequence was 3,600 ms. These homophonic and polyphonic sequences were transposed into six major keys (C major, C# major, D major, D# major, E major, and F major). In the original sequences' (standard) final chord, an out-of-key note was presented independently (single deviance) or simultaneously (double deviance) as tonal deviance at the soprano and bass voices in the soprano–bass deviant condition, at the tenor and bass voices in the tenor–bass deviant condition, and at the alto and tenor

voices in the alto–tenor deviant condition. Here, all possible combinations of outer and inner voices are considered: outer + outer (i.e., soprano–bass), inner + outer (i.e., tenor–bass), and inner + inner (i.e., alto–tenor). This is because outer voices are more salient than inner voices (Huron, 1989) and this difference may confound with the presence of HVSE. Moreover, for the combination of inner and outer voices, the soprano voice was excluded because the soprano is the main melody in homophony and is expected to cause segregation. Thus, the tenor–bass combination was used to investigate the HVSE of the inner + outer voices. Deviant notes were created by altering the original note to the nearest out-of-key note (see Figure 1 for details). Note that the same harmonically irregular chords were used for the homophonic and polyphonic sequences in all conditions.

Because of the sequence length limitation, motifs in polyphonic sequences were restricted to two voice parts. The motifs were created in the soprano and bass voices for the soprano–bass and tenor–bass deviant conditions, and in the alto and tenor voices for the alto–tenor deviant condition. This modification was made to ensure that at least one of the examined voice parts had a motif so that each voice part was perceived as polyphonic. For each condition, three types of polyphonic sequences were composed. The stimulus samples are available at https://osf.io/agux4/.

Procedure

All three conditions were separately conducted as different online experiments following the same procedure. The participants provided written informed consent to participate in the study and provided information about their age, gender, and musical experience. They then adjusted their own acoustic devices (e.g., headphones or speakers) to an optimal sound level. The experimental task was then explained to them,

and they proceeded to the experiment after a practice session. The task was a Go/NoGo task in which the participants were expected to respond with a keypress as quickly and accurately as possible when they detected a deviant, while withholding the response when the standard chord was detected. The experimental instructions were "Your task is to press the space key when either of the following occurs: (1) the higher (highest) note of the last chord is musically incorrect, (2) the lower (lowest) note of the last chord is musically incorrect, either independently or simultaneously. Press the space key as quickly and accurately as possible. If neither (1) nor (2) occurs, do not respond and wait for the next musical sequence to play." The homophonic and polyphonic conditions were presented in separate blocks. Each trial began with the presentation of a fixation cross, followed by the sequence after a 600 ms interval. The fixation cross was terminated by a response or 1,200 ms after the onset of the final chord. The next trial started 500 ms after the offset of the final chord. The homophonic and polyphonic conditions were conducted in a counterbalanced order. Each musical texture condition consisted of three blocks containing 60 trials (10 trials for each deviant chord type and 30 trials for the standard chord). Consequently, the total number of trials was 30 for each deviant trial and 90 for each standard trial. After each block, participants were allowed to rest and were provided with feedback on their performance in the preceding block (hit rate and number of false alarm responses). Before the experimental session, participants separately completed practice trials for the homophonic and polyphonic conditions in different blocks in which all three deviant chords (2 trials each) and the standard chord (6 trials) were randomly presented. Each entire experiment took approximately 40 minutes to complete.

Statistical Analysis for Voice Part Perception

The analysis methods used in all deviant conditions remained consistent. The MRTs were submitted to a two-way repeated measures analysis of variance (ANOVA), with musical texture (homophony and polyphony) and deviance type (double deviant and fastest single deviant) as factors. This analysis aimed to examine the RSE in the homophonic and polyphonic sequences. The single-deviant condition with the shortest average MRT across participants was selected as the fastest single deviant and used in the ANOVA. A Bayesian ANOVA was conducted as an additional analysis to examine the presence or absence of the deviance type effect. When the RSE was observed, the violation of RMI was tested using CDFs of the RTs in all single- and double-deviant conditions (F_S, F_A, F_T, F_B, F_{SB}, F_{TB}, and F_{AT}). The RMI was defined as the inequality, stating that the CDF of the double-deviant condition is less than or equal to the sum of the CDFs of the single-deviant conditions. For instance, the RMI was defined as $F_{SB}(t) + F_C(t) \leq F_S(t) + F_B(t)$ in the soprano–bass deviant condition. Here, $F_C(t)$ was introduced to control the effects of guess responses by incorporating false alarm RTs in the NoGo trials (i.e., kill-the-twin correction: Eriksen, 1988; Ineq. 8: Gondan & Minakata, 2016).

Although the kill-the-twin correction was not preregistered, it is more appropriate for controlling the effect of guess response than trimming the RT to a specific range (as preregistered), because Miller's original RMI does not condition the RT within a specific time range (Miller, 1982). The CDF of the double-deviant condition was compared with the sum of the CDFs of the single-deviant conditions, and the difference was tested using a permutation test. The CDFs of each participant were divided into 10 deciles, and the first 5 decile points were submitted to the permutation test (Gondan, 2010). In the permutation test, $d_i = F_{SB}(t) + F_C(t) - F_S(t) - F_B(t)$

was calculated at each decile point i, and the largest *t* value, t_{max}, was determined among them (one sided: $d_i > 0$). The criterion *t* value for a significant violation of the RMI, t_{crit}, was obtained from the null distribution generated by randomly shuffled data (see details in Gondan, 2010; Gondan & Minakata, 2016). The violation of the RMI was supported when t_{max} exceeded t_{crit}. The significance level was set at .05, and the post-hoc analysis was adjusted using the Bonferroni correction.

Statistical Analysis for High-Voice Superiority Effect

The analysis methods used in all deviant conditions remained consistent. The MRTs of the higher- and lower-deviant conditions were submitted to a two-way repeated measures ANOVA with musical texture and deviance type (higher-voice deviant and lower-voice deviant) as factors. A Bayesian two-way ANOVA was also conducted using the same variable. The hit rates of the higher-voice and lower-voice deviant conditions were submitted to a generalized linear model (GLM) with binomial distribution and logit link using the R package, car library (Fox & Weisberg, 2019). Although this was a change from the preregistered ANOVA, we used the GLM analysis because hit rate data (especially, those with a ceiling effect) may not meet the assumptions of ANOVA. The formula was "hit = musical texture + deviance type + musical texture × condition" and the type II test and the likelihood ratio test were used. To control for overdispersion, variance was compensated by estimating the dispersion parameter using a quasi-binomial. The significance level was set at .05, and the post-hoc analysis was adjusted using the Bonferroni correction.

Results

Voice Part Perception

Figure 2 illustrates the MRT and CDF for each condition. The descriptive statistics and results of the ANOVA are summarized in Table 1. All the results showed that the RSE was consistently observed through the conditions. However, the RMI results were different across the combinations of deviant voices and musical textures.

In the soprano–bass deviant condition, a two-way ANOVA revealed a significant main effect of deviance type (see Table 1 for the detailed statistics). The effect of musical texture and the interaction were not significant, indicating the presence of the RSE, irrespective of the musical textures. The results of the permutation tests showed that F_{SB} was significantly larger than the sum of the single-deviant CDFs within the first to fifth deciles in the polyphonic sequence, $t_{max} = 5.83$, $t_{crit} = 2.19$, $p < .001$, but not in the homophonic sequence, $t_{max} = 1.35$, $t_{crit} = 2.21$, $p = .240$. The RMI was violated only in the polyphonic sequence.

In the tenor–bass deviant condition, the effect of deviance type and the interaction were significant. However, the effect of musical texture was not significant. The post-hoc tests revealed that the MRT of the double deviant was significantly shorter than that of the fastest single deviant (i.e., bass deviant) both in the homophonic and the polyphonic sequences, $ps < .001$, indicating the presence of the RSE, irrespective of musical texture. Moreover, the MRT for the bass deviant was significantly shorter in the homophonic sequence than in the polyphonic sequence, $p = .021$, but not in the double deviant, $p = 1.000$. The results of the permutation tests showed that F_{TB} was significantly larger than the sum of the single-deviant CDFs within the first to fifth deciles in the polyphonic sequence, $t_{max} = 5.65$, $t_{crit} = 2.20$, $p < .001$, but not in the homophonic sequence, $t_{max} = 1.24$, $t_{crit} = 2.22$, $p = .273$. Again, the RMI was violated only in the polyphonic sequence.

In the alto–tenor deviant condition, the main effect of deviance type was significant, indicating the presence of the RSE, irrespective of the musical textures. Although the main effect of musical texture was significant (i.e., RTs were shorter in polyphony than in homophony), the interaction was not significant. The results of the permutation tests showed that F_{AT} was not significantly larger than the sum of the single-deviant CDFs within the first to fifth deciles, either in the polyphonic sequence, $t_{max} = -0.25$, $t_{crit} = 2.18$, $p = .866$, or in the homophonic sequence, $t_{max} = -2.29$, $t_{crit} = 2.16$, $p = .999$. The RMI was not violated in either musical texture in this condition.

High-Voice Superiority Effect

Figure 3 illustrates the MRT and the hit rate. In the soprano–bass and alto–tenor deviant conditions, the HVSE was observed, characterized by shorter MRTs and higher hit rates in the higher voice deviant than the lower voice deviant. However, the HVSE was not observed in the tenor–bass deviant condition.

The top panel of Figure 3 shows the MRT. The descriptive statistics of MRT and the analysis results of MRT are presented in Table 2. In the soprano–bass deviant condition, a two-way ANOVA revealed the significance of deviance type and the interaction. The effect of musical texture was not significant. The post-hoc tests revealed that the MRT of the soprano deviant was shorter than the bass deviant in both the homophonic and polyphonic sequences, suggesting the presence of the HVSE, *ps* < .001. In the tenor–bass deviant condition, the effect of musical texture and deviance type was significant, and the interaction was not significant. Unexpectedly, the MRT of the bass deviant was significantly shorter than that of the tenor deviant in both the homophonic, $p < .001$, and polyphonic sequences, $p = .023$, and the HVSE was not observed. In the alto–tenor deviant condition, the effect of musical texture and deviance

type were significant, indicating the presence of the HVSE. The interaction was not significant. The post-hoc tests revealed that the MRT of the alto deviant was shorter than the tenor deviant in both the homophonic and polyphonic sequences, suggesting the presence of the HVSE, *ps* < .001.

The bottom panel of Figure 3 shows the hit rate. Table 3 shows the descriptive statistics of the hit rate, the false alarm rate, and the results of the GLM. In the soprano–bass deviant condition, the GLM revealed the significance of deviance type, but not musical texture and the interaction. These results suggest that the HVSE was observed in both musical textures. In the tenor–bass deviant condition, the effect of deviance type was significant, but not musical texture and the interaction, suggesting a higher hit rate for the bass deviant. Therefore, the tenor–bass deviant condition did not produce the HVSE in the behavioral response. In the alto–tenor deviant condition, the effects of musical texture and deviance type were significant, but the interaction was not. These results suggest that the hit rate of the alto deviant was higher than that of the tenor deviant, reflecting the presence of the HVSE. The hit rate was significantly higher in the polyphonic sequence than in the homophonic sequence.

Discussion

The present study examined whether separate voice part perception is specific to polyphony by comparing the underlying mechanism of RSEs between the polyphonic and homophonic sequences. The RSEs were observed in all deviant conditions in both the polyphonic and homophonic sequences. These results suggest that each voice part was perceived separately in homophony as well as in polyphony. In the soprano–bass and tenor–bass deviant conditions, the RMI was violated only in the polyphonic

sequence but not in the homophonic sequence. The HVSE was observed in both homophonic and polyphonic sequences in the soprano–bass and tenor–bass deviant conditions, characterized by shorter MRTs and higher hit rates for higher voice deviants than lower voice deviants. However, the HVSE was not observed in the tenor–bass deviant condition of either musical texture.

Voice Part Perception in Homophony and Polyphony

The occurrence of RSEs in both musical textures suggests that separate voice part perception seems to be common in Western tonal music. Both the overall harmony (chord) and the tonality of individual voice parts may be evaluated separately in homophony. Because frequency ranges can be used as a cue to segregate tone sequences (Bregman et al.,1990; Bregman & Campbell, 1971), it is understandable that voice parts in both musical textures were segregated. Nevertheless, evidence of coactivation was observed only in the polyphonic sequence. In RSE studies, violations of RMI have been observed when redundant signals are different enough to be processed as two discrete pieces of information (e.g., different perceptual dimensions: Fiedler et al., 2011; Mordkoff & Danek, 2011; Mordkoff & Yantis, 1993). Based on this property of the RSE, the present results suggest that each voice part is perceived more separately in polyphony than in homophony. In studies of bimodal integration, coactivation results have been interpreted as evidence of the divided attention assuming activations in separate perceptual modules (Miller, 1982) and integrative processing (Schröger & Widmann, 1998). Therefore, the present results of the polyphonic sequence coincide with the integrative model (Bigand et al., 2000), which accounts for the fact that the separate voice part streams are processed in an integrative manner.

In addition to frequency, other musical features such as asynchrony (Huron,

2008; Vos, 1995) and timbre (Cusack & Roberts, 2000; Deike et al., 2004; Oh et al., 2022) can serve as cues for auditory stream segregation. Compared to homophony, polyphony tends to be composed by avoiding synchronous note onsets that cause tonal fusion (Huron, 2008). Synchrony facilitates fused perception and makes it difficult to form segregation (Bregman & Pinker, 1978; Dewitt & Crowder, 1987; Micheyl et al., 2013), whereas asynchrony facilitates perceptual segregation (Vos, 1995). Therefore, in the present study, asynchrony in polyphony may have facilitated the perception of separate voice parts compared to homophony, where voices were presented synchronously. However, in orchestral music and instrumental ensembles, where different voice parts are played by different timbres (i.e., instruments), voice parts may be perceived just as separately in homophony as in polyphony. It would be interesting to investigate whether the perceptual segregation of homophony becomes equivalent to polyphony when different instruments play the voice parts.

One exception to the above-mentioned difference between polyphony and homophony occurred in the combination of inner voice parts (i.e., the alto–tenor deviant condition). In that case, the violation of the RMI was not observed, even in the polyphonic sequence. Previous studies have demonstrated lower perceptual sensitivity for inner voice parts compared to outer voice parts (Huron, 1989; Thompson & Cuddy, 1989). For example, Huron (1989) demonstrated lower detection rates and longer reaction times for entries of inner voice parts compared to outer voice parts when participants monitored the polyphony with changing numbers of voice parts in real time. Because of the reduced perceptual sensitivity for inner voice parts, the segregation of inner voice parts may be difficult, and perceptual separability may be attenuated even in the polyphonic sequence.

Separate tonal processing of each voice part in homophonic sequences may be attributed to the voice leading in Western harmony. Through a theoretical review, Huron (2001) suggested that the voice leading assists the segregation of each voice part in accordance with the perceptual grouping principles. The horizontal perception (e.g., perception of melodic lines) and vertical perception (e.g., perception of harmony) can function in parallel during music listening. In the homophonic sequence of the present study, voice leading served as a cue, and each voice part may have been perceived as a different auditory stream. Previous studies have shown that harmony and voice leading are closely related to harmonic expectancies (Poulin-Charronnat et al., 2005; Wall et al., 2020). Thus, harmonic expectation through cadential motion (Bigand et al. 1996; Janata, 1995) may have confounded the present results and the voice parts containing the V–I progression may have heightened harmonic expectations and attracted the attention to the voice line. In addition, the inclusion of out-of-key notes as deviant stimuli could be a cue for the separation of voice parts (e.g., mistuning tone in harmonics: Alain et al., 2001). Future research could investigate whether voice leading is one of the cues used to separate voice parts in music. Nevertheless, it should be emphasized that the current results of the homophonic and polyphonic sequences are comparable in that the same target chords appeared at the same end position in both sequences.

HVSE Reflected in Behavioral Responses

The shorter MRT and higher hit rates in higher voice parts were observed not only in the soprano–bass deviant condition but also in the alto–tenor deviant condition. Previous neurophysiological studies have examined the HVSE using polyphonic sequences consisting of two melodies (Fujioka et al., 2005, 2008; Huberth & Fujioka,

2017; Marie et al., 2012). The HVSE results in the present study expand the HVSE findings in the highest voice part to the higher voice part, although this effect may be limited to a specific pitch register, as shown by Trainor et al. (2014). Moreover, the HVSE may not be specific to polyphony, and this effect may be general in processing Western tonal music, because the HVSE was also observed in the homophonic sequence.

In comparing the tenor and bass voice parts, the HVSE was not observed in either the polyphonic or homophonic sequences. This result has two possible explanations. First, the HVSE may be limited to a sufficiently high-pitched register. Second, the salience of outer voice parts may have interfered with the salience of the higher voice part, because the tenor–bass deviant condition was a combination of inner voice and outer voice parts. Thompson and Cuddy (1989) demonstrated that detection accuracy in the detection task of key change was higher in the outer voice than in the inner voice. In line with this, the saliency of the outer bass voice part may have been higher than that of the inner tenor voice part in the present study.

The detection performance was higher for the homophonic sequence than for the polyphonic sequence in the deviant conditions that included the bass deviant, while reverse patterns were observed in the combination of the inner voice part deviants. Specifically, the MRT for the bass deviant was shorter in the homophonic sequence than in the polyphonic sequence. These results suggest that the bass voice part was more strongly perceived in the homophony, where chord progressions were clearer than in the polyphony. This interpretation aligns with that of Schwitzgebel and White (2021), who proposed that harmonic expectations are sensitively affected by bass patterns as well as pitch-class content of chord, as evidenced by higher ratings of conclusiveness in a chord

progression with root position chords (paradigmatic bass pitches) compared to a chord progression with inverted chords (nonparadigmatic bass pitches). In contrast, the inner voice parts were easily detectable in polyphony, where several voice parts were more independent than in homophony.

Limitations

The present study has three possible limitations. First, the response competition may have inhibited the Go response in the single-deviant conditions (Eriksen & Eriksen, 1974; Grice et al., 1984). Previous studies have shown that a non-signal (NoGo) channel inhibits the response to a target signal (Go) when the signal channel coexists with the non-signal condition. However, coactivation was observed only in polyphony but not in homophony, even when exactly the same target chord was used in both musical textures. Therefore, while the redundancy gains observed in the present study may be partly attributed to response competition, these gains cannot be fully explained by response competition alone.

Second, the instructions in the present study may have encouraged divided attention to each voice part. The participants were told to detect deviants in each voice part. Thus, the RSE in the homophonic sequence can be attributed to the facilitation of separate voice part perception due to the instructions. Although the separate voice part perception could be facilitated by the instructions, greater separation of voice parts was still evident in polyphony compared to homophony. Future research should aim to replicate these findings by altering the instructions to focus on the deviance of the entire chord object rather than on each individual voice part.

Third, the polyphonic sequences used in the present study had only two motifs. In the soprano–bass and alto–tenor conditions, both target parts had motifs, while only

the bass part had a motif in the tenor–bass deviant condition. Because the RSEs and evidence of coactivation were observed regardless of the presence of motifs, the number of motifs did not affect the perception of separate voice parts in the present study. However, the absence of the HVSE in the tenor–bass deviant condition may be attributed to the lack of a motif in the tenor voice part in the tenor–bass deviant condition. The salience of the higher tenor voice part may have been weaker than that of the bass voice part, which had a motif. Nevertheless, it should be noted that the conclusion of the present study, that relatively higher pitch produces the HVSE, was evidenced by the occurrence of the HVSE in two comparable conditions (i.e., the soprano–bass and alto–tenor deviant conditions), where motifs were present in two voice parts. Although this study did not cover all possible combinations of voice parts, it may be interesting to consider all possible combinations systematically. Future research may find interactive effects between the HVSE, pitch-heigh, voice position (outer or inner), and the presence of the motifs.

Conclusion

The present study compared voice part perception in polyphonic and homophonic sequences using the RSE paradigm to investigate whether separate voice part perception is specific to polyphony. The presence of RSEs in both sequences indicates separate voice part perception regardless of musical texture. The violation of RMI observed only in the polyphonic sequence suggests a greater separation of voice parts in polyphony compared to homophony. However, the inner voice parts remained less separable, even in polyphony. Additionally, the higher voice part was perceived more saliently than the lower voice part in both musical textures, while the bass deviant

was detected more quickly than the tenor deviant in the tenor–bass deviant condition. The privilege in polyphony may be an enhanced segregation of voice parts, while the privilege in homophony may be a weighting perception of the bass voice part to facilitate the evaluation of harmony.

References

Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(5), 1072–1089. https://doi.org/10.1037/0096-1523.27.5.1072

Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2020). Power contours: Optimizing sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*, *26*(3), 295–314. https://doi.org/10.1037/met0000337

Barrett, K. C., Ashley, R., Strait, D. L., Skoe, E., Limb, C. J., & Kraus, N. (2021). Multi-voiced music bypasses attentional limitations in the brain. *Frontiers in Neuroscience*, *15*, 588914. https://doi.org/10.3389/fnins.2021.588914

Von Békésy, G. (1960). *Experiments in hearing*. Mcgraw Hill.

Bigand, E., McAdams, S., & Forêt, S. (2000). Divided attention in music. *International Journal of Psychology*, *35*(6), 270–278. https://doi.org/10.1080/002075900750047987

Bigand, E., Parncutt, R., & Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics*, *58*(1), 125–141. https://doi.org/10.3758/BF03205482

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. The MIT Press.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, *89*(2), 244–249. https://doi.org/10.1037/h0031163

Bregman, A. S., Liao, C., & Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian Journal of Psychology*, *44*(3), 400–413. https://doi.org/10.1037/h0084255

Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, *32*(1), 19–31. https://doi.org/10.1037/h0081664

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics*, *62*(5), 1112–1120. https://doi.org/10.3758/BF03212092

Dewitt, L. A., & Crowder, R. G. (1987). Tonal fusion of consonant musical intervals: The oomph in stumpf. *Perception & Psychophysics*, *41*(1), 73–84. https://doi.org/10.3758/BF03208216

Deike, S., Gaschler-Markefski, B., Brechmann, A., & Scheich, H. (2004). Auditory stream segregation relying on timbre involves left auditory cortex. *NeuroReport*, *15*(9), 1511–1514. https://doi.org/10.1097/01.wnr.0000132919.12990.34

Disbergen, N. R., Valente, G., Formisano, E., & Zatorre, R. J. (2018). Assessing top-down and bottom-up contributions to auditory stream segregation and integration with polyphonic music. *Frontiers in Neuroscience*, *12*, 121. https://doi.org/10.3389/fnins.2018.00121

Dowling, W. J. (1973). The perception of interleaved melodies. *Cognitive Psychology*, *5*(3), 322–337. https://doi.org/10.1016/0010-0285(73)90040-6

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.1007/978-3-319-57111-9_9085

Eriksen, C. W. (1988). A source of error in attempts to distinguish coactivation from separate activation in the perception of redundant targets. *Perception & Psychophysics*, *44*(2), 191–193. https://doi.org/10.1111/j.1545-5300.1977.00363.x

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Fiedler, A., Schröter, H., & Ulrich, R. (2011). Coactive processing of dimensionally redundant targets within the auditory modality? *Experimental Psychology*, *58*(1), 50–54. https://doi.org/10.1027/1618-3169/a000065

Fox, J., & Weisberg, S. (2019). An R Companion to Applied Regression, Third edition, Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Fujioka, T., Trainor, L. J., & Ross, B. (2008). Simultaneous pitches are encoded separately in auditory cortex: An MMNm study. *NeuroReport*, *19*(3), 361–366. https://doi.org/10.1097/WNR.0b013e3282f51d91

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *Journal of Cognitive Neuroscience*, *17*(10), 1578–1592. https://doi.org/10.1162/089892905774597263

Gondan, M. (2010). A permutation test for the race model inequality. *Behavior Research Methods*, *42*(1), 23–28. https://doi.org/10.3758/BRM.42.1.23

Gondan, M., & Minakata, K. (2016). A tutorial on testing the race model inequality. *Attention, Perception, and Psychophysics*, *78*(3), 723–735. https://doi.org/10.3758/s13414-015-1018-y

Gregory, A. H. (1990). Listening to polyphonic music. *Psychology of Music*, *18*(2), 163–170. https://doi.org/10.1177/0305735690182005

Grice, G. R., Canham, L., & Boroughs, J. M. (1984). Combination rule for redundant information in reaction time tasks with divided attention. *Perception & Psychophysics*, *35*(5), 451–463. https://doi.org/10.3758/BF03203922

Hausfeld, L., Disbergen, N. R., Valente, G., Zatorre, R. J., & Formisano, E. (2021). Modulating cortical instrument representations during auditory stream segregation and integration with polyphonic music. *Frontiers in Neuroscience*, *15*, 635937. https://doi.org/10.3389/fnins.2021.635937

Hove, M. J., Marie, C., Bruce, I. C., & Trainor, L. J. (2014). Superior time perception for lower musical pitch explains why bass-ranged instruments lay down musical rhythms. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(28), 10383–10388. https://doi.org/10.1073/pnas.1402039111

Huberth, M., & Fujioka, T. (2017). Neural representation of a melodic motif: Effects of polyphonic contexts. *Brain and Cognition*, *111*, 144–155. https://doi.org/10.1016/j.bandc.2016.11.003

Huron, D. (1989). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception*, *6*(4), 361–382. https://doi.org/10.2307/40285438

Huron, D. (1991). Tonal consonance versus tonal fusion in polyphonic sonorities. *Music Perception*, *9*(2), 135–154. https://doi.org/10.2307/40285526

Huron, D. (2001). Tone and Voice: A Derivation of the Rules of Voice-Leading from Perceptual Principles. *Music Perception*, *19*(1), 1–64. https://doi.org/10.1525/mp.2001.19.1.1

Huron, D. (2008). Asynchronous preparation of tonally fused intervals in polyphonic. *Empirical Musicology Review*, *3*(1), 69–72. https://doi.org/10.18061/1811/31695

Huron, D., & Fantini, D. A. (1989). The avoidance of inner-voice entries: Perceptual evidence and musical practice. *Music Perception*, *7*(1), 43–47. https://doi.org/10.2307/40285447

Ishida, K., & Nittono, H. (2023). Multidimensional regularity processing in music: An examination using redundant signals effect. *Research Square*, https://doi.org/10.21203/rs.3.rs-3226380/v1

Janata, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts in music. *Journal of Cognitive Neuroscience*, *7*(2), 153–164. https://doi.org/10.1162/jocn.1995.7.2.153

Janata, P., Tillmann, B., & Bharucha, J. J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognitive, Affective and Behavioral Neuroscience*, *2*(2), 121–140. https://doi.org/10.3758/CABN.2.2.121

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York, NY: Oxford University Press.

Marie, C., Fujioka, T., Herrington, L., & Trainor, L. J. (2012). The high-voice superiority effect in polyphonic music is influenced by experience: A comparison of musicians who play soprano-range compared with bass-range instruments.

Psychomusicology: Music, Mind, and Brain, *22*(2), 97–104. https://doi.org/10.1037/a0030858

Micheyl, C., Hanson, C., Demany, L., Shamma, S., & Oxenham, A. J. (2013). Auditory stream segregation for alternating and synchronous tones. Journal of Experimental Psychology: Human Perception and Performance, 39(6), 1568. https://psycnet.apa.org/doi/10.1037/a0032241

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247–279. https://doi.org/10.1016/0010-0285(82)90010-X

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics*, *40*(5), 331–343. https://doi.org/10.3758/BF03203025

Miller, J. (2004). Exaggerated redundancy gain in the split brain: A hemispheric coactivation account. *Cognitive Psychology*, *49*(2), 118–154. https://doi.org/10.1016/j.cogpsych.2003.12.003

Miller, J., & Ulrich, R. (2003). Simple reaction time and statistical facilitation: A parallel grains model. *Cognitive Psychology*, *46*(2), 101–151. https://doi.org/10.1016/S0010-0285(02)00517-0

Mordkoff, J. T., & Danek, R. H. (2011). Dividing attention between color and shape revisited: Redundant targets coactivate only when parts of the same perceptual object. *Attention, Perception, and Psychophysics*, *73*(1), 103–112. https://doi.org/10.3758/s13414-010-0025-2

Mordkoff, J. T., & Yantis, S. (1993). Dividing attention between color and shape: Evidence of coactivation. *Perception & Psychophysics*, *53*(4), 357–366. https://doi.org/10.3758/BF03206778

Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology*, *42*(1), 25–32. https://doi.org/10.1111/j.1469-8986.2005.00256.x

Oh, Y., Zuwala, J. C., Salvagno, C. M., & Tilbrook, G. A. (2022). The Impact of Pitch and Timbre Cues on Auditory Grouping and Stream Segregation. *Frontiers in Neuroscience*, *15*, 725093. https://doi.org/10.3389/fnins.2021.725093

Poulin-Charronnat, B., Bigand, E., & Madurell, F. (2005). The influence of voice leading on harmonic priming. *Music Perception*, *22*(4), 613–627. https://doi.org/10.1525/mp.2005.22.4.613

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*(5), 574–590. https://doi.org/10.1111/j.2164-0947.1962.tb01433.x

Saupe, K., Koelsch, S., & Rübsamen, R. (2010). Spatial selective attention in a complex auditory environment such as polyphonic music. *The Journal of the Acoustical Society of America*, *127*(1), 472–480. https://doi.org/10.1121/1.3271422

Schröger, E., & Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology*, *35*(6), 755–759. https://doi.org/10.1017/S0048577298980714

Schröter, H., Ulrich, R., & Miller, J. (2007). Effects of redundant auditory stimuli on reaction time. *Psychonomic Bulletin and Review*, *14*(1), 39–44. https://doi.org/10.3758/BF03194025

Schwitzgebel, E., & White, C. W. (2021). Effects of chord inversion and bass patterns on harmonic expectancy in musicians. *Music Perception: An Interdisciplinary Journal*, *39*(1), 41–62. https://doi.org/10.1525/mp.2021.39.1.41

Sloboda, J., & Edworthy, J. (1981). Attending to two melodies at once: The of key relatedness. *Psychology of Music*, *9*(1), 39–43. https://doi.org/10.1177/03057356810090010701

Smyth, G., Hu, Y., Dunn, P., Phipson, B., Chen, Y., & Smyth, M. G. (2017). Package 'statmod'. R Documentation. Package for R programming version 1.5.0. http://CRAN.R-project.org/package=statmod

Thompson, W. F., & Cuddy, L. L. (1989). Sensitivity to key change in chorale sequences: A comparison of single voices and four-voice harmony. *Music Perception*, *7*(2), 151–168. https://doi.org/10.2307/40285455

Trainor, L. J., Marie, C., Bruce, I. C., & Bidelman, G. M. (2014). Explaining the high voice superiority effect in polyphonic music: Evidence from cortical evoked potentials and peripheral auditory models. *Hearing Research*, *308*, 60–70. https://doi.org/10.1016/j.heares.2013.07.014

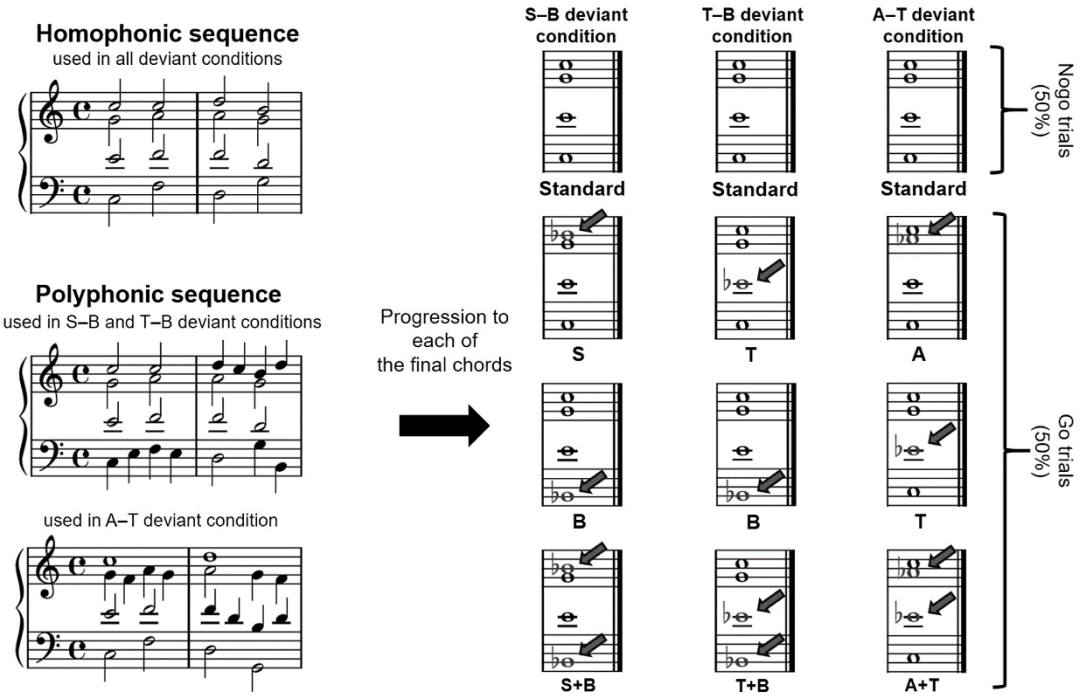
Vliegen, J., Moore, B. C. J., & Oxenham, A. J. (1999). The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *The Journal of the Acoustical Society of America*, *106*(2), 938–945. https://doi.org/10.1121/1.427140

Vos, J. (1995). Perceptual separation of simultaneous complex tones: The effect of slightly asynchronous onsets. *Acta Acustica*, *3*(5), 405–416. https://doi.org/10.1121/1.401500

Wall, L., Lieck, R., Neuwirth, M., & Rohrmeier, M. (2020). The impact of voice leading and harmony on musical expectancy. *Scientific Reports*, *10*(1), 5933. https://doi.org/10.1038/s41598-020-61645-4

Figure 1.

Homophonic and polyphonic sequences used in the current study



Note. Sequences of each musical texture were followed by a final chord with or without deviant notes (i.e., Go or NoGo trials). S, A, T, and B indicate soprano deviant, alto deviant, tenor deviant, and bass deviant, respectively. The notes of the deviant voice parts are indicated by arrows. Both Go and NoGo trials were presented with equal probability. Note that the same final chords followed homophonic and polyphonic sequences. For polyphony, soprano and bass motifs were used in the S–B and T–B deviant conditions, whereas alto and tenor motifs were used in the alto–tenor deviant condition to ensure that at least one deviant voice part had a motif.

Table 1.

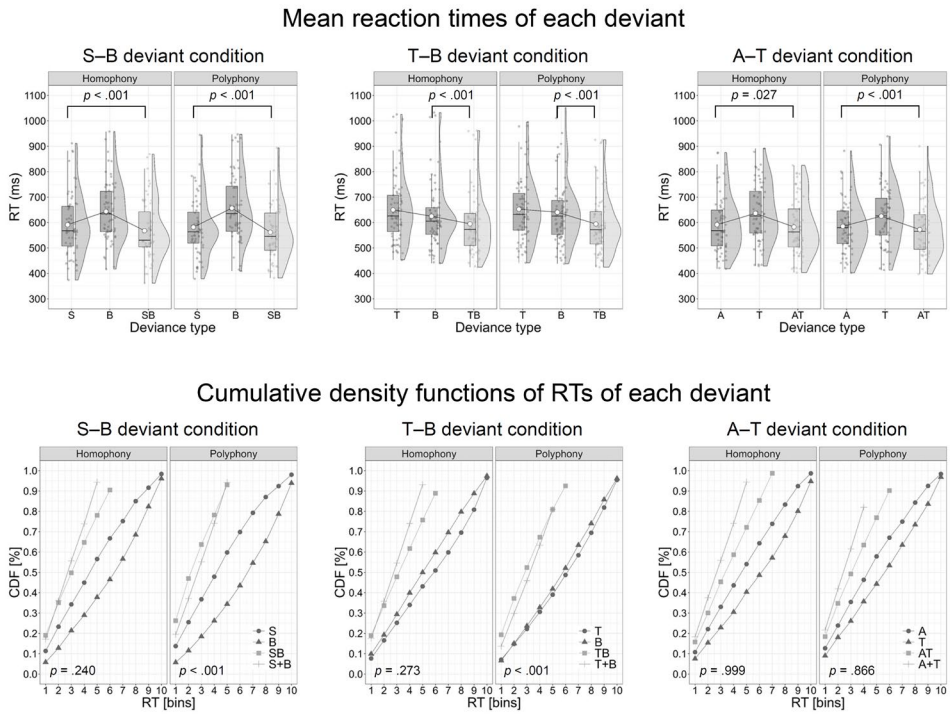
Mean and standard deviation (SD) of MRTs for each deviant condition and summary of statistical analysis

	Soprano–Bass (*n* = 64)						Tenor–Bass (*n* = 76)						Alto–Tenor (*n* = 68)					
	Homophony			Polyphony			Homophony			Polyphony			Homophony			Polyphony		
	S	B	SB	S	B	SB	T	B	TB	T	B	TB	A	T	AT	A	T	AT
M	591	644	568	582	657	562	650	625	595	653	641	594	592	636	583	585	625	572
SD	120	117	117	118	129	110	127	124	122	124	128	117	107	115	109	102	111	107
	Musical texture		Deviance type		Interaction		Musical texture		Deviance type		Interaction		Musical texture		Deviance type		Interaction	
F	3.54		58.14		0.42		2.71		104.58		9.95		4.70		11.96		1.15	
df	1, 63		1, 63		1, 63		1, 75		1, 75		1, 75		1, 67		1, 67		1, 67	
p	.065		< .001		.519		.104		< .001		.002		.034		< .001		.288	
η^2	.053		.480		.007		.035		.582		.117		.066		.151		.017	
BF_{10}	1.01		1.82×10^{7}		0.23		0.66		8.19×10^{12}		12.84		1.73		29.89		0.30	

Note. S, A, T, and B indicate the soprano deviant, alto deviant, tenor deviant, and bass deviant, respectively. The bottom panel shows the results of a two-way ANOVA with musical texture and deviance type as factors. In the row of the degrees of freedom (*df*), the left and right values indicate the numerator and denominator *df*, respectively.

Figure 2.

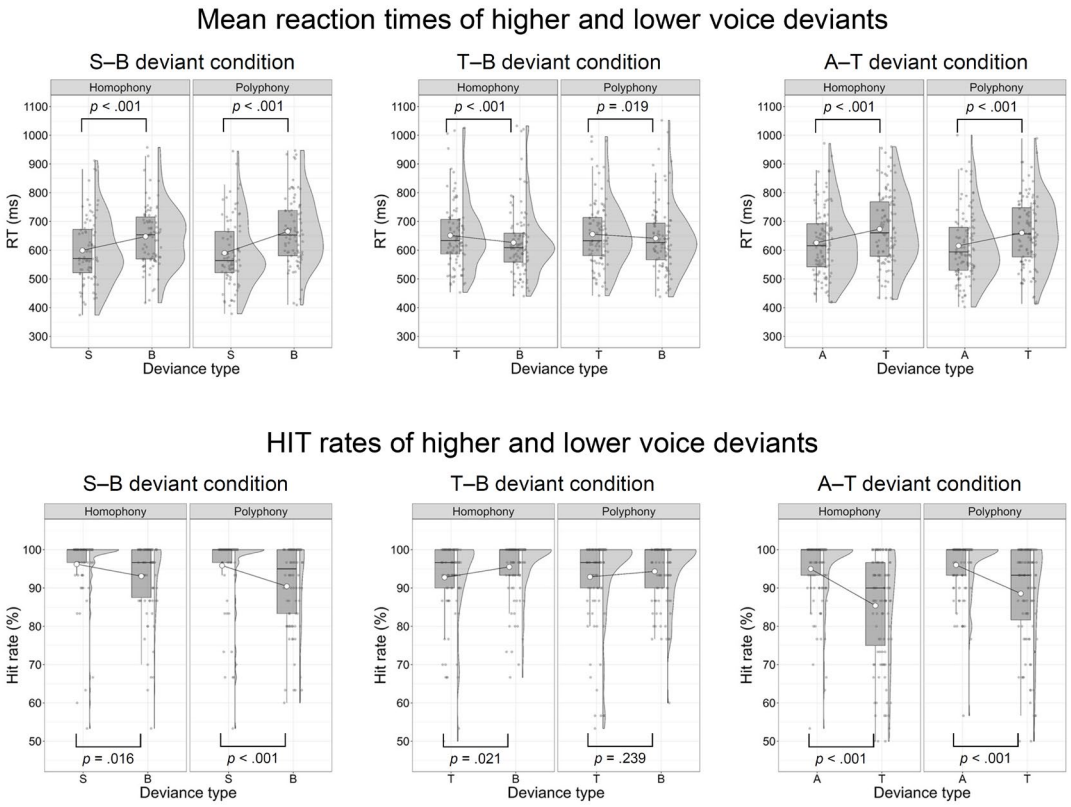
MRTs and CDFs for single and double deviants



Note. The top panel shows the MRTs for single and double deviants. S, A, T, and B indicate the soprano deviant, alto deviant, tenor deviant, and bass deviant, respectively. The white dots indicate the average MRTs. The significance lines indicate the comparison of the fastest single deviants and the double deviants. The bottom panel shows the CDFs of the RTs for each deviant condition. The horizontal axis indicates bins of RTs, with RTs arranged in order of decreasing time and separated into 10 deciles. The vertical axis indicates the cumulative probability. The purple CDF, calculated as single deviant + single deviant, exceeds 1.0 and is thus truncated before reaching 1.0. The *p* values written in the bottom left of each CDF plot indicate the results of the RMI permutation tests.

Figure 3.

MRTs and hits for higher and lower voice deviants



Note. The top panel shows the MRTs for higher (blue) and lower (red) voice deviants. S, A, T, and B indicate the soprano deviant, alto deviant, tenor deviant, and bass deviant, respectively. The white dots indicate the average MRTs. The bottom panel shows the hit rates of the higher and lower voice deviants. The white dots indicate the average hit rates.

Table 2.

Mean and standard deviations (SD) of MRTs for each deviant condition used in analysis of the HVSE and summary of statistical analysis

	Soprano–Bass (n = 82)						Tenor–Bass (n = 99)						Alto–Tenor (n = 111)					
	Homophony			Polyphony			Homophony			Polyphony			Homophony			Polyphony		
	S	B	SB	S	B	SB	T	B	TB	T	B	TB	A	T	AT	A	T	AT
M	599	648	570	590	665	565	651	626	593	656	642	598	626	673	618	616	661	605
SD	116	106	110	118	119	104	118	116	112	115	118	109	120	125	120	117	119	119
	Musical texture		Deviance type		Interaction		Musical texture		Deviance type		Interaction		Musical texture		Deviance type		Interaction	
F	0.90		73.73		14.77		5.32		12.62		3.29		7.90		91.70		0.22	
df	1, 81		1, 81		1, 81		1, 98		1, 98		1, 98		1, 110		1, 110		1, 110	
p	.346		< .001		< .001		.023		< .001		.073		.006		< .001		.638	
η^2	.011		.477		.154		.051		.114		.032		.067		.455		.002	
BF_{10}	0.22		1.43×10^{10}		110.63		1.68		45.06		0.69		5.80		1.63×10^{13}		0.16	

Note. S, A, T, and B indicate the soprano deviant, alto deviant, tenor deviant, and bass deviant, respectively. The bottom panel shows the results of a two-way ANOVA with musical texture and deviance type as factors. In the row of degrees of freedom (*df*), the left and right values indicate the numerator and denominator *df*, respectively.

Table 3.

Mean and standard deviations (SD) of hit and false alarm (FA) rates for each deviant condition used in the HSVE analysis

	Soprano–Bass (*n* = 82)								Tenor–Bass (*n* = 99)								Alto–Tenor (*n* = 111)							
	Homophony				Polyphony				Homophony				Polyphony				Homophony				Polyphony			
	S	B	SB	FA	S	B	SB	FA	T	B	TB	FA	T	B	TB	FA	A	T	AT	FA	A	T	AT	FA
M	96.2	93.1	98.9	2.7	95.9	90.4	99.5	2.0	92.8	95.5	99.1	2.9	92.9	94.3	99.0	2.5	95.0	85.4	95.1	4.1	96.0	88.6	97.0	3.9
SD	9.0	10.0	3.0	5.1	9.7	11.2	1.5	3.4	10.6	7.5	2.5	5.1	11.6	8.0	3.2	6.1	8.7	14.6	8.0	5.6	7.2	13.4	5.7	6.7

	Musical texture	Deviance type	Interaction	Musical texture	Deviance type	Interaction	Musical texture	Deviance type	Interaction
χ^2	1.72	14.00	0.35	0.30	4.93	0.63	3.86	65.28	0.02
p	.190	< .001	.551	.583	.026	.429	.049	< .001	.893

Note. S, A, T, and B indicate the soprano deviant, alto deviant, tenor deviant, and bass deviant, respectively. The bottom panel shows the results of a GLM analysis with musical texture and deviance type as factors.

Supplementary Materials

Supplementary Table S1.

Detailed attributes of the participants used for analysis of the RSE and HVSE

	Analysis of the RSE									
	sample size	age		sex		handedness			musical experience	
		M	*SD*	Man	Woman	Left	Right	ambidextrous	*M*	*SD*
Soprano–Bass	64	41.5	6.4	42	22	4	58	2	4.4	7.0
Tenor–Bass	76	43.3	9.1	42	34	3	71	2	6.8	8.6
Alto–Tenor	68	41.8	9.8	30	38	5	60	3	6.5	8.3
	Analysis of the HVSE									
	sample size	age		sex		handedness			musical experience	
		M	*SD*	Man	Woman	Left	Right	ambidextrous	*M*	*SD*
Soprano–Bass	82	43.1	7.6	56	26	5	74	3	3.9	6.4
Tenor–Bass	99	42.5	9.6	59	40	4	92	3	5.8	7.9
Alto–Tenor	111	42.7	9.9	63	48	7	101	3	4.9	7.6

Note. Musical experience was assessed by asking about years of extracurricular musical training.

Simulation of Null Hypothesis Rejection Rates

To validate the planned sample size (58) indicated by the power contour, we conducted a post-hoc simulation of null hypothesis rejection rates under a coactivation model. Based on the mean RTs and SDs of our previous study's data (single deviant 1: $M = 583$, $SD = 118$; single deviant 2: $M = 572$, $SD = 98$; double deviant: $M = 514$; $SD = 99$), we randomly generated the RTs of each participant following the inverse Gaussian distribution using "statmod" (Smyth et al., 2017, version 1.5.0), which is an R package. Inverse Gaussian distribution was selected because the RT does not take a minus value. As with the permutation test reported in the main text, the first five decile points were submitted to the permutation test. When the sample size was 58, the number of deviant trials was 24, and the iteration was 10,000, the null hypothesis rejection rate was 99.4%. This result indicates that the sample size of the present study ($N \geq 64$) was large enough to detect a violation of RMI.

Re-analysis of the RSE by comparing the shortest single-deviant MRT of each participant and the double-deviant MRT

The single-deviant MRT was defined as the fastest MRT of each participant and compared to the double-deviant MRT. In the soprano–bass deviant condition, the means (*SDs*) of the single-deviant MRTs were 585 (117) ms and 580 (118) ms for homophony and polyphony, respectively. A two-way ANOVA revealed a significant main effect of the deviance type (see Supplementary Table S2 for detailed statistics). The effect of musical texture and the interaction were not significant, indicating the presence of the RSE, irrespective of musical textures.

In the tenor–bass deviant condition, the means (*SDs*) of the single-deviant

MRTs were 612 (125) ms and 628 (121) ms for homophony and polyphony, respectively. A two-way ANOVA revealed a significant main effect of deviance type. The effect of deviance type and the interaction were significant, but the effect of musical texture was not. Post-hoc tests revealed that the RSE occurred in both musical textures (homophony: $p < .001$; polyphony: $p < .001$) and that the single-deviant MRT was faster in homophony than in polyphony ($p = .002$).

In the alto–tenor deviant condition, the means (*SDs*) of the single-deviant MRTs were 588 (105) ms and 581 (102) ms for homophony and polyphony, respectively. A two-way ANOVA revealed a significant main effect of deviance type. The effects of musical texture and deviance type were significant, but the interaction was not. These results suggest that the RSE occurred in both musical textures and the MRT was shorter in polyphony than in homophony. These results were virtually the same as those reported in the main text: RSEs were observed in all conditions.

Supplementary Table S2.

Summary of statistical results of RSE using the shortest single-deviant MRT of each participant

	Soprano–Bass (n = 64)			Tenor–Bass (n = 76)			Alto–Tenor (n = 67)		
F	1.92	48.48	0.11	3.25	78.72	17.02	4.97	6.04	1.08
df	1, 63	1, 63	1, 63	1, 75	1, 75	1, 75	1, 67	1, 67	1, 67
p	.171	< .001	.744	.076	< .001	< .001	.029	.017	.302
η^2	.029	.435	.002	.041	.512	.185	.069	.083	.016
BF_{10}	0.51	1.15×10^6	0.20	0.91	2.48×10^{10}	206.40	1.91	2.47	0.30

Note. The table shows the results of a two-way ANOVA with musical texture and deviance type as factors. These results are virtually the same as those reported in the main text (i.e., RSEs were present for all conditions in homophony and polyphony). In the row of degrees of freedom (*df*), the left and right values indicate the numerator and denominator *df*, respectively.

Comparison of tenor-deviant MRTs between the alto–tenor and tenor–bass conditions

To test whether the salience of voice parts differed between the presence and absence of the motif, the tenor-deviant MRTs in the alto–tenor (with motif) and tenor-bass (without motif) conditions were compared. A Welch *t*-test failed to detect a difference between the two conditions for both homophony, $t(142) = -0.67$, $p = .502$, Cohen's $d = -0.11$, $BF_{10} = 0.22$, and polyphony, $t(142) = -1.40$, $p = .163$, Cohen's $d = -0.23$, $BF_{10} = 0.44$. From the current data, the effect of the presence of the motif was inconclusive.

Re-analysis of the HVSE using the participant samples for the RSE analysis

The HVSE was reanalyzed using the participant samples for the RSE analysis because the exclusion criterion of the HVSE data was different from that of the RSE data. As in the preregistered analysis, the MRTs were submitted to the 2-way ANOVA with musical texture (homophony and polyphony) and deviance type (higher-voice deviant and lower-voice deviant). As in the main text, the hit rates were submitted to the GLM (binomial distribution and logit link) with musical texture and deviance type. All results are summarized in Supplementary Tables S3–4. The results were almost identical to those reported in the main text.

Supplementary Table S3.

Summary of statistical analysis of the HVSE based on the participant samples for the RSE analysis

	Soprano–Bass (n = 64)			Tenor–Bass (n = 76)			Alto–Tenor (n = 67)		
F	0.15	70.78	8.66	3.64	12.71	3.76	3.31	73.47	0.84
df	1, 63	1, 63	1, 63	1, 75	1, 75	1, 75	1, 67	1, 67	1, 67
p	.700	< .001	.005	.060	< .001	.056	.073	< .001	.363
η^2	.002	.529	.121	.046	.145	.048	.047	.523	.012
BF_{10}	0.18	1.35×10^9	9.25	0.78	44.85	0.95	1.00	4.22×10^9	0.24

Note. The table shows the results of a two-way ANOVA with musical texture and deviance type as factors. The post-hoc tests for the interaction in the soprano–bass condition showed that the MRT was shorter for the soprano deviant in both the homophony and polyphony. These results are virtually the same as the analysis reported in the main text. In the row of degrees of freedom (*df*), the left and right values indicate the numerator and denominator *df*, respectively.

Supplementary Table S4.

Mean and standard deviations (SD) of hit and false alarm (FA) rates for each deviant condition based on the participant samples for the RSE analysis

	Soprano–Bass (*n* = 64)								Tenor–Bass (*n* = 76)								Alto–Tenor (*n* = 67)							
	Homophony				Polyphony				Homophony				Polyphony				Homophony				Polyphony			
	S	B	SB	FA	S	B	SB	FA	T	B	TB	FA	T	B	TB	FA	A	T	AT	FA	A	T	AT	FA
M	98.4	96.1	99.2	2.0	98.5	93.9	99.8	1.7	96.4	97.4	99.4	1.5	96.7	96.2	99.3	1.0	98.7	94.8	98.2	2.3	98.8	96.9	99.3	1.9
SD	3.5	5.8	2.7	3.2	4.0	7.5	0.7	3.2	4.4	4.8	1.9	1.9	5.2	5.8	2.4	1.5	3.4	5.6	3.3	3.5	3.5	4.5	1.8	3.2

	Musical texture	Deviance type	Interaction	Musical texture	Deviance type	Interaction	Musical texture	Deviance type	Interaction
χ^2	2.26	27.23	1.26	0.74	0.17	1.60	4.01	29.10	0.85
p	.133	< .001	.261	.391	.682	.206	.045	< .001	.357

Note. S, A, T, and B indicate the soprano deviant, alto deviant, tenor deviant, and bass deviant, respectively. The bottom panel shows the results of a GLM analysis with musical texture and deviance type as factors. These results are almost identical to those reported in the main text: the presence of the HVSE in the soprano–bass and alto-tenor conditions, but not in the tenor–bass condition.