TITLE

Frontotemporal connectivity changes reflect figure–ground reversal in multivoiced music

Chan Hee Kim[1, 2], Jeong-Eun Seo[3], and Chun Kee Chung[1, 2, 4, 5, *]

[1] Interdisciplinary Program in Neuroscience, Seoul National University College of Natural Science, Seoul, Korea,

[2] MEG Center, Department of Neurosurgery, Seoul National University Hospital, Seoul, Korea,

[3] Department of Musicology, Seoul National University, Seoul, Korea,

[4] Department of Brain and Cognitive Science, Seoul National University College of Natural Science, Seoul, Korea,

[5] Department of Neurosurgery, Seoul National University Hospital, Seoul, Korea

* Correspondence: chungc@snu.ac.kr

ABSTRACT

The neural processes underlying the perception of prominently heard voices (figure) and less prominently heard voices (ground), particularly when this prominence is naturally reversed while listening to music, remain inadequately investigated. We investigated changes in frontotemporal connectivity in response to figure–ground reversal within multivoiced music using Mozart's variation KV 265, including the "Twinkle, Twinkle, Little Star (TTLS)" melody based on our previous study of frontotemporal connectivity (TTLS connectivity), influenced by the emergence of the TTLS melody in the upper voice at the beginning of each variation, and independent of lower voices. We examined the consistency of the TTLS connectivity pattern across repetitions of the same phrases within each variation. Notably, TTLS connectivity changed only for the final repetition between variations sharing the same TTLS melody. This result indicates a perceptual reversal, processing the TTLS melody initially perceived as figure into the ground. Our data effectively illustrate how the brain dissects voices within the multidimensional structures of continuously changing music, reconstructing them through a momentary switch in figure–ground processes. These findings provide valuable insights into the neural processes associated with the dynamic experience of listening to music, a ubiquitous aspect of human life.

KEYWORDS

Multivoiced music; Figure–ground reversal; Heschl's gyrus; Inferior frontal gyrus; Effective connectivity; Magnetoencephalography

INTRODUCTION

Humans can perceive multivoiced streams in the sounds of nature, language, and music, both separately and together. When listening to polyphonic music, distinct voices with varying pitch ranges can be perceived independently. The melodic line of the highest voice tends to stand out perceptually due to its pitch range. However, the perceptual dominance of the melodic line is not fixed and can be influenced by the listener's attention and choice. Indeed, when listening to music, individuals can choose to fucus on listening and attending to any specific voice within the musical structure. The present study aimed to capture the moment when the human brain's recognition of the melodic line shifts to the lower voice. This shift reflects the dynamic experience of listening to music.

The concepts of figure and ground in Gestalt psychology[1,2] find parallels in the perceptually prominent and less prominent voices in music. Similar to visual stimuli (Fig. 1B), where attention can shift freely between a white star and a black ground, attention can shift to different voices within music (Fig. 1A). The auditory stream distinguishes the upper voice through the higher pitches[3,4], but the lower voice can also be perceived as a figure, depending on its prominence in the musical context[5]. Predicting how listener's attention shifts between voices is possible by analyzing the formal structure of the music. The figure–ground relationship for the upper and lower voices can change in terms of the repetition of passages[6]. In later repetitions, the upper voice may be recognized as the ground, and the lower voice as the figure. Musical compositions, including symphonies, sonatas, variations, and rondos, contain formal structures. Musical structure, which encompasses elements like pitch, tonality, and harmonic structure, is implicitly learned through musical experience[7]. Recognizing similarities in music helps developing hypotheses, categorizing inputs, and processing variables[8]. For example, the song "Twinkle, Twinkle, Little Star" follows the ternary form of A (a + a) + B (b + a') + B (b + a'). Understanding this structure enables the recognition of repeated passages and the prediction of subsequent sections while listening to the piece. Anticipation of repeated passages of a familiar melody may naturally help shift attention toward the other voices. Accordingly, the structure including the figure and ground may be implicitly predictable in the brain of an individual who already heard "Twinkle, Twinkle, Little Star (TTLS)".

In our recent study using Mozart's variation KV 265[9], we observed changes in frontotemporal connectivity from the left Heschl's gyrus (HG) to the left inferior frontal gyrus (IFG). This connectivity was contingent upon whether the first five notes of the original melody of TTLS — "C5-C5-G5-G5-A5" — were present in the initial passages of each variation. Frontotemporal connectivity responded to the presence or absence of the TTLS melody, indicating its recognition as a figure (as illustrated in Set 1 of Fig. 1B, showing two distinct patterns based on the presence of the inner star). These findings were consistent across target phrases lasting 2.1 s, equivalent to a five-quarter note

length of “C5-C5-G5-G5-A5” at the beginning of each variation in two pairs — “*Variations I* and *II*” and “*Variations III* and *IV*”. In our previous study[9], we ruled out the effect of repetition[10] by analyzing only the “C5-C5-G5-G5-A5” in the beginning of each condition. Although both pairs shared the same presence or absence of the original melody, they differed in the lower voices (Fig. 2). This finding indicates that frontotemporal connectivity, specifically from the left IFG to the right HG, is sensitive not to the lower voices with different rhythmic patterns (see also the descriptions of rhythmic variation in Fig. 2) but specifically to the TTLS melody. If the relationship between figure and ground reverses after repetitions of the target phrases, the strength of connectivity consistent between variations with the TTLS melody undergo changes.

We wanted to investigate the possibility of differences in the frontotemporal connectivity related to the TTLS melody-based process within variations sharing the same TTLS melody. Our goal was to capture moments when the relationship between figure and ground switched in voices, moving momentarily from the upper voice of the TTLS melody to the lower voice, as the two patterns were distinct based on the outer patterns in Set 2 of Fig. 1B. Mozart’s variation KV 265 is an ideal stimulus to test changes in figure–ground perception and its reversal in the target phrases, including the TTLS melody appearing repeatedly in each variation. Figure–ground reversal can be verified when those figures (melodies) are the same, which make the neural responses are also consistent. In Mozart’s variation KV265, the two conditions were the *Variation II* and *Variation IV* (Fig. 2D). The target phrase containing the TTLS melody “C5-C5-G5-G5-A5,” was repeated four times per variation (Fig. 2). Our previous findings[9] showed that frontotemporal connectivity patterns were consistent between *Variation II* and *Variation IV*, with the same TTLS melody and the different lower voices. These results were measured for the 1st target phrase (T1 in Fig. 2D), considering the effect of repetition. However, target phrases with the same TTLS melody appear repeatedly in *Variations II* and *IV*. Our current aim was to identify instances of figure–ground reversal experienced by subjects while listening to voices within the Mozart’s variation KV 265, using frontotemporal connectivity data from the same magnetoencephalography (MEG) dataset determining TTLS connectivity. In this study, we tested the frontotemporal connectivity between *Variations II* and *IV* for all target phrases (T1 to T4) and whether a change between the *Variations II* and *IV* is observed in target phrases, in the formal structure of music (Fig. 2). We hypothesized that given a change between *Variations II* and *IV* in the frontotemporal connectivity observed among a target phrase, unlike in our previous results[4], it would be caused by figure–ground reversal. Consequently, we developed two hypotheses: 1) If the frontotemporal connectivity for target phrases does not significantly differ between *Variations II* and *IV*, the same TTLS melody is processed as the figure, with the lower voice remaining as the ground. 2) If there is a significant difference in frontotemporal connectivity between the two variations, the

differences in the lower voices lead to a conversion of the lower voices to the figure, converting the upper voices into the ground.

We estimated the effective connectivity from the left IFG to the right HG using linearized time-delayed mutual information (LTDMI)[11], a measure employed in our previous study[9]. LTDMI allows for estimating information transmission between two regional sources and observing the directionality in interhemispheric and interregional connectivity. We anticipated that the perceptual reversal between the figure of the melodic line and the ground of the lower voice line would elicit distinct responses within the repetitions of the target phrases in the musical structure (Fig. 2B). The dynamic interplay of brain networks between bilateral IFGs and HGs is instrumental in processing various musical components. Particularly, the left IFG plays a critical role in implicit learning of syntax in both language and music[12], also contributing to memory contolling[13] and attention[14]. Thus, the directionality of connection from the left IFG to the right HG efficiently explains the memory-based processing of input stimuli, including the aforementioned processes. This approach allowed us to explore how changes in effective connectivity within these regions correlate with perceptual shifts associated with repeated target phrases, shedding light on the intricate dynamics of melodic and lower voice processing.

While the central concept of this study revolves the shifting of attention between voices, the primary objective of this study was not to investigate connectivity changes through the deliberate manipulation of subjects' attention to a particular voice or to assess the effects of attention. Instead, this study aimed to discern the musical scene in which the figure and the ground switch within the brain signals of individuals listening to the naturalistic stimuli of Mozart's variation KV 265. During the MEG experiment, all subjects passively listened to the musical stimulus once, without any instructions to focus their attention on a specific voice or melody. Furthermore, we did not collect behavioral responses regarding attention to the voices after the MEG experiment. Our inquiries were based on the careful selection of target phrases capable of eliciting figure–ground reversal within a formal structure based on music theory. In addition, we constrained our hypotheses by the TTLS connectivity, which effectively controlled for other musical dynamics outside of the TTLS melody.

RESULTS

Differences in frontotemporal connectivity

Frontotemporal connectivity from the left IFG to the right HG was assessed using LTDMI[9]. We independently performed a Wilcoxon signed-rank test for each target period to confirm that the change in the figure–ground relationship between the two voices could be reflected in the frontotemporal connectivity from the left IFG to the right HG across variations. The difference in

frontotemporal connectivity from the left IFG to the right HG between *Variations II* and *IV* was calculated for four target phrases (T1–T4), each repeated four times per variation (Fig. 2). The difference between the two variations was significant only in T4 among all target phrases, as indicated by the Wilcoxon signed-rank test ($Z = -2.112$, $P = 0.035$; see Fig. 3A). In T4, frontotemporal connectivity was enhanced in *Variation IV* compared with *Variation II*. However, significant differences were not observed in T1–T3 ($P > 0.05$ in all cases). The significance level (α) for the null hypothesis was independently tested for the individual target periods (T1–T4), since individual target phrases existed in completely different musical contexts before and after in the formal structure (Fig. 2B). In addition, we confirmed that, among 12 connections between the bilateral IFGs and HGs, the only significant result corresponded specially to frontotemporal connectivity from the left IFG to the right HG (Supplementary Table 1).

DISCUSSION

There were differences in frontotemporal connectivity from the left IFG to the right HG between the two variations, specifically in the final target phrase of T4, without differences in the preceding three phrases. According to our hypothesis, an examination of the disparity in LTDMI between the two variations suggests that the TTLS melody of the upper voice underwent a reversal, becoming the ground in T4, but remained the figure in T1–T3 (Fig. 3B). This indicates that subjects may have been particularly attentive to the TTLS melody up to T3, but in T4, a natural shift occurred, leading them to focus more on the lower voice due to habituation from repeated listening. Each variation included cue phrases like “Up above the world...,” predicting the recurrence of the TTLS melody (see also Fig. 2B and Supplementary Fig. 1). T4 was introduced after the second iteration of the cue phrase (Fig. 2B), helping subjects anticipate the reappearance of the TTLS melody. The predictability of the upper voice through repetition might have facilitated participants’ recognition of the lower voice[6]. In T4, the frontotemporal connectivity for *Variation IV* surpassed that for *Variation II*. Considering the role of frontotemporal connectivity linked to the TTLS melody-based process, this heightened connectivity might indicate a response to novel information regarding passages containing the TTLS melody[9], reflecting the processing of the figure–ground reversal towards lower voices.

Top-down processing of a familiar melody can significantly influence the figure–ground relationship[15,16]. The left IFG is crucial in processing familiarity[17] and plays a key role in music-syntactic processes, indicating implicit learning[12] and contributing to memory retrieval[13,18]. In addition, the left IFG is involved in differentiating between melody and accompaniment[19]. On the other hand, the right HG predominantly processes tone deviance [20,21] and the segregation of auditory streams[22]. The right auditory cortex has been the dominant site for music processing[23]. Accordingly, the directional information flow in frontotemporal connectivity could explain how the left IFG and the

right HG collaborate in processing target phrases. The IFG and HG are pivotal areas for music perception, and their connectivity is discussed in relation to syntax processes[24-26], categorization[27], and the working memory of melody process[28]. Thus, frontotemporal connectivity between the IFG and HG is likely modulated by both a top-down process based on knowledge of the TTLS melody and a bottom-up process based on new information on voices accumulated while sequentially listening to target and cue phrases in each variation[29,30].

Familiarity is critical for explaining the figure–ground experiment[15,31,32]. The target phrases in our stimuli also involved the TTLS song, used in studies[33-36] related to perception of familiar melodies. Familiarity would be naturally implied in the theme and all variations, considering that Mozart's variation KV265 is based on a theme of the TTLS melody. Logically, the familiarity implied in the TTLS melody is influential on figure–ground perception and its reversal. However, in our previous study focusing on the TTLS melody, we could not directly prove the effect of familiarity on frontotemporal connectivity. Frontotemporal connectivity was the connectivity that changed with the presence and absence of the TTLS melody in our previous study[9]. In our present study, the TTLS melody appearing repeatedly was the same in *Variation II* and *Variation IV*. The effects of familiarity are consistent with both *Variations II* and *IV*. TTLS connectivity was applied to prove the figure–ground reversal in our present study, throwing the following results: (1) Consistent TTLS connectivity between *Variations II* and *IV* from T1 to T3 indicates that the TTLS melody in the target phrases served as the figure. (2) Inconsistent TTLS connectivity between *Variations II* and *IV* in T4 indicates that the TTLS melody in the target phrases functioned as the ground, while the lower voice became the figure.

Research on figure–ground reversal originated in visual studies, notably using the classic Rubin vase-faces model[16,37]. Monkey data have highlighted the involvement of the primary visual cortex in neuronal activities in response to figure–ground perception of visual stimuli[38-41]. Visual figure–ground discrimination as well as behavior, has been explored in various animals such as mice[42], flies[43], and honey bees[44]. Notably, figure–ground processing in visual objects differs between anesthetized and awake animals[45]. In the present study, behavioral responses to differences between T1, T2, T3, and T4 were not measured, because task performance while listening to music did not align with the objective of this study. Brain responses were solely measured in the context of figure–ground reversal using real music, without controlling for attention. While previous studies have selectively manipulated stimuli or controlled subjects' attention to focus on specific auditory streams[19,46-51], attention's critical role in figure–ground perception has been acknowledged[52]. Despite our results, it remains unclear whether both figure and ground were processed attentively or pre-attentively, given the absence of intentional attention control in our study. Artificially manipulated stimuli, while useful in some cases[53,54], deviate from the complex and dynamic nature of real musical experiences. Music listeners

can automatically process information such as syntactic errors and tone deviations even without intentional attention[55,56]. Although listeners may focus on a particular voice while listening to music, the ever-changing flow of music can result in such brief perceptible moments that the figure–ground relationship continuously shifts without listeners consciously realizing. Our results successfully captured the moment when the figure–ground relationship between the upper and lower voices switched, as evidenced by the different frontotemporal connectivity from the IFG to the right HG for repeated phrases in the two variations.

For auditory stimuli, figure–ground perception is not exclusive to humans; it has also been observed in animals, in studies on figure–ground processes investigated through noise and sound. Extracellular recordings demonstrated increased multiunit activities in the auditory area[57]. Further, auditory areas have been the focus of previous human studies on stream segregation through selective attention and figure–ground segregation[46,47,49,51]. In fact, numerous studies on auditory figure–ground segregation were based on the concept of auditory scene analysis[58] and grouping[1], using tasks that discriminate a sound pattern as a figure from a ground of tone and chord sequences with irregularities in spectral and temporal properties[53,54,59]. Regions involved in the discrimination of figure–ground segregation reported in fMRI and EEG studies[53,54,59] comprise the primary auditory area, superior temporal sulcus, superior temporal gyrus, intraparietal sulcus, medial/superior frontal gyrus, and cingulate cortex. In addition to these findings, the processing of multiple voices in this study involved the IFG and the HG, both of which exhibited information transmission. This difference in involvement may be attributed to the nature of our stimuli, which comprised real music containing contextual richness rather than edited notes or sequences. Furthermore, the participants in this study passively listened to Mozart's variation KV 265 without being asked to concentrate selectively on specific musical parts. The involvement of the left IFG and right HG might reflect the entire process of naturally grasping, comparing, and understanding the voices of target phrases in the context, rather than simply listening to a sound. Therefore, information transmission in frontotemporal connectivity from the left IFG to the right HG could encompass all processes of discriminating target phrases with the TTLS melody repeated within the real music of Mozart's variation KV 265. This includes segregating the lower voice from a stream containing the TTLS melody and perceiving the figure–ground reversal of the TTLS melody and its lower voices in this context.

We decided to use a musical composition by Wolfgang Amadeus Mozart rather than artificially composed stimuli. Our attempt to use a real music had several limitations. Our study paradigm was examined in a single trial. This result may be limited compared with those obtained through repeated measurements using artificially composed stimuli, even if our attempt reflects a ubiquitous human experience. With respect to the frontotemporal connectivity from the left IFG to the right HG, we could not determine which specific (novel) information about musical features within the two

variations triggered the observed connectivity enhancement and attenuation. Moreover, in this study, musicians with greater sensitivity to voice perception and superior voice perception ability[4,50] were not recruited as participants. Music training affects selective attention[48]. Although nonmusicians can separately perceive the figure from the ground[53], future studies should verify these results in musicians. In addition, this study concentrated on two regions, the left IFG and the right HG. Our findings should be verified at the whole-brain level. Nonetheless, using Mozart's variation KV 265 was a compelling and valuable endeavor in unraveling the fundamental neural processes associated with processing real music featuring multiple voices and understanding the human experience of music. Our data effectively illustrated how the brain can dissect voices from the multidimensional structures of continuously changing music and reconstruct them through a momentary switch in figure–ground processes.

MATERIALS AND METHODS

Participants

Participants comprised 25 healthy individuals, all nonmusicians, 15 women and 10 men with a mean age of 26.8 ± 3.4 years old. None had received formal musical training. All participants were right-handed, with a mean Edinburg Handedness coefficient of 95.7 ± 7.1. The study adhered to the principles of the Declaration of Helsinki and received approval from the Institutional Review Board of the Clinical Research Institute at Seoul National University Hospital (IRB No. C-1003-015-311). The research procedures were conducted in accordance with relevant guidelines and regulations, and all subjects provided informed consent, adhering to ethical guidelines.

Stimuli

Mozart's KV 265 consists of the theme and 12 variations on "Ah! vous dirai-je Maman". This study focused on *Variation II* and *Variation IV*, with the same TTLS melody. *Variations II* and *IV* share the same TTLS melody but differs in the lower parts, transforming through rhythmic changes to 8th note triplets and 16th notes (semiquavers), respectively. The tonality and harmonic structure are the same for both variations (Fig. 2A). Each variation was based on the ternary form of A (a + a) + B (b + a') + B (b + a'). Phrases including the "C5-C5-G5-G5-A5" melody were repeated four times in each variation (Fig. 2B).

MEG recording

In a magnetically shielded room, the participants listened to Mozart's KV 265 while watching a silent movie clip (Love Actually, 2003, Universal Pictures, USA) for approximately 5 min. Musical stimuli were generated using STIM^{2TM} (Neuroscan, Charlotte, NC, USA) and presented binaurally at 100 dB through MEG-compatible earphones (Tip-300, Nicolet, Madison, WI, USA). MEG signals were

recorded using a 306-channel whole-head MEG system (Elekta Neuromag Vector ViewTM, Helsinki, Finland) with a sampling frequency of 1,000 Hz and a bandpass filter of 0.1–200 Hz.

MEG analysis

Source analysis

The MEG source signals of epochs from −100 ms to 2,100 ms after the onset of each condition for four regional sources of bilateral HGs and IFGs, bandpass filtered at 14–30 Hz, were extracted using BESA 5.1.8.10 (MEGIS Software GmbH, Gräfelfing, Germany) after electrooculograms, electrocardiograms, and muscle artifacts were removed. Standard Talairach coordinates (x, y, and z in millimeters) for bilateral HGs (transverse, BA 41, BA 42) and IFGs (triangular part, BA 45) across participants were adapted from previous research[9]. These coordinates were as follows: left HG (−53.5, −30.5, and 12.6), right HG (55.4, −30.5, and 12.6), left IFG (−55.5, 11.7, and 20.6), and right IFG (53.5, 12.7, and 20.6).

Time window

Variations II and *IV* were chosen as the two conditions for estimating connectivity differences, including the identical TTLS melody. Time windows were 2,100 ms long, incorporating the "C5-C5-G5-G5-A5" melody, as based on a previous study[9]. Time windows of 2,100 ms appeared four times per variation (Fig. 2D and Supplementary Fig. 1). The lower voices accompanied by the "C5-C5-G5-G5-A5" melody differed between the two variations (Fig. 2). There was only one trial for each target phrase because the stimulus was real music. Each individual subject's data included the data of four target phrases for the two variations.

LTDMI analysis

While our primary focus was the regional connection from the left IFG to the right HG, we verified our results for all connections among the bilateral IFGs and HGs. The effective connectivity for the 12 connections between regional sources of the bilateral HGs and IFGs was estimated using MATLAB 7.7.0.471 (Math Works Inc., Natick, MA, USA) (see also Supplementary Table 1 for individual LTDMI values calculated for 12 connections). LTDMI was used to estimate information transmission between two time series based on mutual information (MI)[11,24]. For each subject, the mean LTDMI for the 2,100-ms epoch was calculated for each of 4 target phrases (T1, T2, T3, and T4) × 2 variations (*Variations II* and *IV).* Statistical comparisons of mean LTDMI values for *Variations II* and *IV* were performed using SPSS 21.0 software (IBM, Armonk, NY, USA). Considering different contexts in each target phrase, we conducted the non-parametric *Wilcoxon signed-rank test* due to the non-Gaussian distribution of LTDMI data. In each case, the significance level (α) for rejecting the null

hypothesis (H0, indicating no difference between *Variation II* and *Variation IV* in the mean LTDMI values), was 0.05 (N = 25).

Availability of Data and Materials

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

REFERENCES

1 Wagemans, J. *et al.* A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychol Bull* **138**, 1172-1217, doi:10.1037/a0029333 (2012).

2 Köhler, W. Gestalt psychology. *Psychologische Forschung* **31**, XVIII-XXX (1967).

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

4 Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R. & Pantev, C. Automatic encoding of polyphonic melodies in musicians and nonmusicians. *Journal of cognitive neuroscience* **17**, 1578-1592, doi:10.1162/089892905774597263 (2005).

5 Deutch, D. *Chapter 3. The Perceptual Organization of Streams of Sound*. (Oxford University Press, 2019).

6 Taher, C., Rusch, R. & McAdams, S. Effects of repetition on attention in two-part counterpoint. *Music Perception: An Interdisciplinary Journal* **33**, 306-318 (2016).

7 Tillmann, B., Bharucha, J. J. & Bigand, E. Implicit learning of tonality: a self-organizing approach. *Psychological review* **107**, 885 (2000).

8 Narmour, E. Music expectation by cognitive rule-mapping. *Music Perception* **17**, 329-398 (2000).

9 Kim, C. H. *et al.* Increased fronto-temporal connectivity by modified melody in real music. *PloS one* **15**, e0235770 (2020).

10 Guo, S. & Koelsch, S. The effects of supervised learning on event-related potential correlates of music-syntactic processing. *Brain research* **1626**, 232-246, doi:10.1016/j.brainres.2015.01.046 (2015).

11 Jin, S. H., Lin, P. & Hallett, M. Linear and nonlinear information flow based on time-delayed mutual information method and its application to corticomuscular interaction. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology* **121**, 392-401, doi:10.1016/j.clinph.2009.09.033 (2010).

12 Sammler, D., Koelsch, S. & Friederici, A. D. Are left fronto-temporal brain areas a prerequisite for normal music-syntactic processing? *Cortex; a journal devoted to the study of the nervous system and behavior* **47**, 659-673, doi:10.1016/j.cortex.2010.04.007 (2011).

13 Watanabe, T., Yagishita, S. & Kikyo, H. Memory of music: roles of right hippocampus and left inferior frontal gyrus. *NeuroImage* **39**, 483-491, doi:10.1016/j.neuroimage.2007.08.024 (2008).

14 Hampshire, A., Chamberlain, S. R., Monti, M. M., Duncan, J. & Owen, A. M. The role of the right inferior frontal gyrus: inhibition and attentional control. *NeuroImage* **50**, 1313-1319 (2010).

15 Nelson, R. A. & Palmer, S. E. Familiar shapes attract attention in figure-ground displays. *Perception & psychophysics* **69**, 382-392 (2007).

16 Strüber, D. & Stadler, M. Differences in top—down influences on the reversal rate of different categories of reversible figures. *Perception* **28**, 1185-1196 (1999).

17 Plailly, J., Tillmann, B. & Royet, J. P. The feeling of familiarity of music and odors: the same neural signature? *Cereb Cortex* **17**, 2650-2658, doi:10.1093/cercor/bhl173 (2007).
18 Platel, H., Baron, J. C., Desgranges, B., Bernard, F. & Eustache, F. Semantic and episodic memory of music are subserved by distinct neural networks. *NeuroImage* **20**, 244-256 (2003).
19 Spada, D., Verga, L., Iadanza, A., Tettamanti, M. & Perani, D. The auditory scene: an fMRI study on melody and accompaniment in professional pianists. *NeuroImage* **102 Pt 2**, 764-775, doi:10.1016/j.neuroimage.2014.08.036 (2014).
20 Sabri, M., Liebenthal, E., Waldron, E. J., Medler, D. A. & Binder, J. R. Attentional modulation in the detection of irrelevant deviance: a simultaneous ERP/fMRI study. *Journal of cognitive neuroscience* **18**, 689-700, doi:10.1162/jocn.2006.18.5.689 (2006).
21 Nan, Y. & Friederici, A. D. Differential roles of right temporal cortex and Broca's area in pitch processing: evidence from music and Mandarin. *Human brain mapping* **34**, 2045-2054, doi:10.1002/hbm.22046 (2013).
22 Snyder, J. S., Alain, C. & Picton, T. W. Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of cognitive neuroscience* **18**, 1-13, doi:10.1162/089892906775250021 (2006).
23 Perani, D. *et al.* Functional specializations for music processing in the human newborn brain. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 4758-4763, doi:10.1073/pnas.0909074107 (2010).
24 Kim, C. H. *et al.* Dissociation of Connectivity for Syntactic Irregularity and Perceptual Ambiguity in Musical Chord Stimuli. *Frontiers in neuroscience* **15**, 693629, doi:10.3389/fnins.2021.693629 (2021).
25 Kim, C. H. *et al.* Change in left inferior frontal connectivity with less unexpected harmonic cadence by musical expertise. *PloS one* **14**, e0223283, doi:10.1371/journal.pone.0223283 (2019).
26 Papoutsi, M., Stamatakis, E. A., Griffiths, J., Marslen-Wilson, W. D. & Tyler, L. K. Is left fronto-temporal connectivity essential for syntax? Effective connectivity, tractography and performance in left-hemisphere damaged patients. *NeuroImage* **58**, 656-664, doi:10.1016/j.neuroimage.2011.06.036 (2011).
27 Roswandowitz, C., Swanborough, H. & Fruhholz, S. Categorizing human vocal signals depends on an integrated auditory-frontal cortical network. *Human brain mapping* **42**, 1503-1517, doi:10.1002/hbm.25309 (2021).
28 Burunat, I., Alluri, V., Toiviainen, P., Numminen, J. & Brattico, E. Dynamics of brain activity underlying working memory for music in a naturalistic condition. *Cortex; a journal devoted to the study of the nervous system and behavior* **57**, 254-269, doi:10.1016/j.cortex.2014.04.012 (2014).
29 Alho, K., Salmi, J., Koistinen, S., Salonen, O. & Rinne, T. Top-down controlled and bottom-up triggered orienting of auditory attention to pitch activate overlapping brain networks. *Brain research* **1626**, 136-145, doi:10.1016/j.brainres.2014.12.050 (2015).
30 Dzafic, I. *et al.* Stronger Top-Down and Weaker Bottom-Up Frontotemporal Connections During Sensory Learning Are Associated With Severity of Psychotic Phenomena. *Schizophrenia bulletin* **47**, 1039-1047, doi:10.1093/schbul/sbaa188 (2021).
31 Palmer, S. E. *Vision science: Photons to phenomenology*. (MIT press, 1999).
32 Hulleman, J. & Humphreys, G. W. A new cue to figure–ground coding: Top–bottom polarity. *Vision research* **44**, 2779-2791 (2004).
33 Trehub, S. E., Morrongiello, B. A. & Thorpe, L. A. Children's perception of familiar melodies: The role of intervals, contour, and key. *Psychomusicology: A Journal of Research in Music Cognition* **5**, 39 (1985).
34 Creel, S. C. The familiar-melody advantage in auditory perceptual development: Parallels between spoken language acquisition and general auditory perception. *Attention, Perception, & Psychophysics* **81**, 948-957 (2019).
35 Besson, M., Faïta, F. & Requin, J. Brain waves associated with musical incongruities differ for musicians and non-musicians. *Neuroscience letters* **168**, 101-105 (1994).
36 Upitis, R. Children's invented notations of familiar and unfamiliar melodies. *Psychomusicology: A Journal of Research in Music Cognition* **9**, 89 (1990).

37 Peterson, M. A., Harvey, E. M. & Weidenbacher, H. J. Shape recognition contributions to figure-ground reversal: Which route counts? *Journal of Experimental Psychology: Human Perception and Performance* **17**, 1075 (1991).
38 Lamme, V. A. f. The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of neuroscience* **15**, 1605-1615 (1995).
39 Super, H., van der Togt, C., Spekreijse, H. & Lamme, V. A. Internal state of monkey primary visual cortex (V1) predicts figure–ground perception. *Journal of Neuroscience* **23**, 3407-3414 (2003).
40 Zhang, N. R. & Von Der Heydt, R. Analysis of the context integration mechanisms underlying figure–ground organization in the visual cortex. *Journal of Neuroscience* **30**, 6482-6496 (2010).
41 Von der Heydt, R. Figure–ground organization and the emergence of proto-objects in the visual cortex. *Frontiers in psychology* **6**, 1695 (2015).
42 Kirchberger, L. *et al.* The essential role of recurrent processing for figure-ground perception in mice. *Science advances* **7**, eabe1833 (2021).
43 Reichardt, W. & Poggio, T. Figure-ground discrimination by relative movement in the visual system of the fly. *Biological cybernetics* **35**, 81-100 (1979).
44 Srinivasan, M. V., Lehrer, M. & Horridge, G. A. Visual figure–ground discrimination in the honeybee: the role of motion parallax at boundaries. *Proceedings of the Royal Society of London. B. Biological Sciences* **238**, 331-350 (1990).
45 Lamme, V. A., Zipser, K. & Spekreijse, H. Figure-ground activity in primary visual cortex is suppressed by anesthesia. *Proceedings of the National Academy of Sciences* **95**, 3263-3268 (1998).
46 Uhlig, M., Fairhurst, M. T. & Keller, P. E. The importance of integration and top-down salience when listening to complex multi-part musical stimuli. *NeuroImage* **77**, 52-61, doi:10.1016/j.neuroimage.2013.03.051 (2013).
47 Barrett, K. C. *et al.* Multi-Voiced Music Bypasses Attentional Limitations in the Brain. *Frontiers in neuroscience* **15**, 588914, doi:10.3389/fnins.2021.588914 (2021).
48 Puschmann, S., Baillet, S. & Zatorre, R. J. Musicians at the Cocktail Party: Neural Substrates of Musical Training During Selective Listening in Multispeaker Situations. *Cereb Cortex* **29**, 3253-3265, doi:10.1093/cercor/bhy193 (2019).
49 Hausfeld, L., Riecke, L., Valente, G. & Formisano, E. Cortical tracking of multiple streams outside the focus of attention in naturalistic auditory scenes. *NeuroImage* **181**, 617-626, doi:10.1016/j.neuroimage.2018.07.052 (2018).
50 Strait, D. L., Slater, J., O'Connell, S. & Kraus, N. Music training relates to the development of neural mechanisms of selective auditory attention. *Developmental cognitive neuroscience* **12**, 94-104, doi:10.1016/j.dcn.2015.01.001 (2015).
51 Ragert, M., Fairhurst, M. T. & Keller, P. E. Segregation and integration of auditory streams when listening to multi-part music. *PloS one* **9**, e84085, doi:10.1371/journal.pone.0084085 (2014).
52 Poort, J. *et al.* The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron* **75**, 143-156 (2012).
53 Toth, B. *et al.* EEG signatures accompanying auditory figure-ground segregation. *NeuroImage* **141**, 108-119, doi:10.1016/j.neuroimage.2016.07.028 (2016).
54 Teki, S., Chait, M., Kumar, S., von Kriegstein, K. & Griffiths, T. D. Brain bases for auditory stimulus-driven figure-ground segregation. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **31**, 164-171, doi:10.1523/JNEUROSCI.3788-10.2011 (2011).
55 Maess, B., Koelsch, S., Gunter, T. C. & Friederici, A. D. Musical syntax is processed in Broca's area: an MEG study. *Nature neuroscience* **4**, 540-545, doi:10.1038/87502 (2001).
56 Naatanen, R., Paavilainen, P., Rinne, T. & Alho, K. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology* **118**, 2544-2590, doi:10.1016/j.clinph.2007.04.026 (2007).
57 Schneider, F. *et al.* Neuronal figure-ground responses in primate primary auditory cortex. *Cell reports* **35**, 109242, doi:10.1016/j.celrep.2021.109242 (2021).

58 Bregman, A. S. *Auditory scene analysis: The perceptual organization of sound*. (MIT press, 1994).
59 O'Sullivan, J. A., Shamma, S. A. & Lalor, E. C. Evidence for Neural Computations of Temporal Coherence in an Auditory Scene and Their Enhancement during Active Listening. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **35**, 7256-7263, doi:10.1523/JNEUROSCI.4973-14.2015 (2015).
60 Bigand, E., McAdams, S. & Forêt, S. Divided attention in music. *International Journal of Psychology* **35**, 270-278, doi:10.1080/002075900750047987 (2000).

ACKNOWLEDGMENT

We sincerely appreciate Ji Hyang Nam for technically supporting MEG data acquisition.

AUTHOR CONTRIBUTIONS

C.H.K.: conceptualization, funding, formal analysis, visualization, writing of original draft, and review and editing of the manuscript; J. S.: advice about music theory, and review and editing of the manuscript; C.K.C.; supervision, funding, and review and editing of the manuscript.

FUNDING

This research was supported by Samsung Research Funding & Incubation Center for Future Technology (SRFC-IT1902-08, Decoding Inner Music Using Electrocorticography), and the Basic Science Research Program through the National Research Foundation of Korea, funded by the Ministry of Science & ICT (NRF-2021R1A4A200180312) and the Ministry of Education (NRF-2022R1I1A1A01073800).

COMPETING INTERESTS

The authors declare no competing interests.

FIGURES

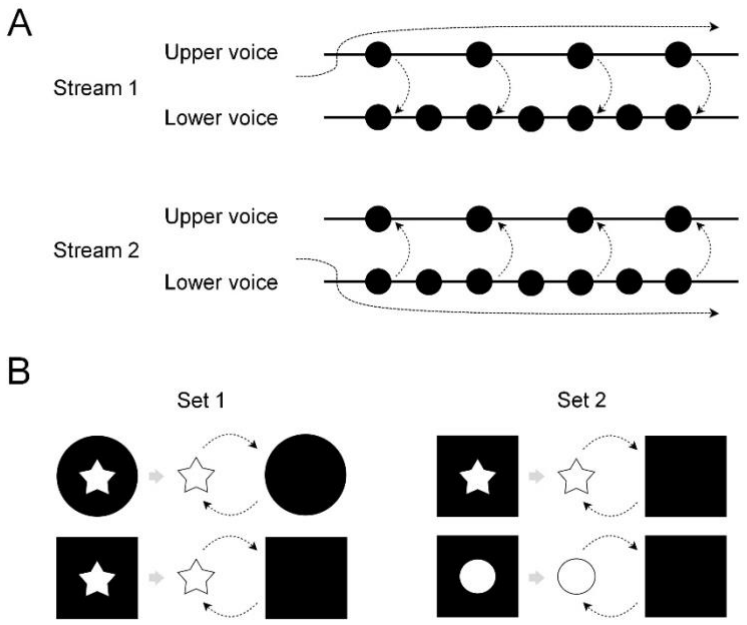

Fig. 1. Figure–ground perception in auditory and visual stimuli. A. In multivoiced music (Stream 1), the lower voice serves as ground while attention is directed to the upper voice as figure. Conversely, in Stream 2, the roles are reversed. This model was adapted from Bigand et al., 2000[60]. B. Set 1 illustrates the figure–ground perception in visual stimuli. When the white stars form the figure perceptually, the black circle and square serve as the ground, and vice versa. Notably, the difference between the two objects lies in the shapes of the black circle and square. In Set 2, both objects share the same black square background. However, the black squares can become figures, whereas the white star and circle are perceived as the ground.

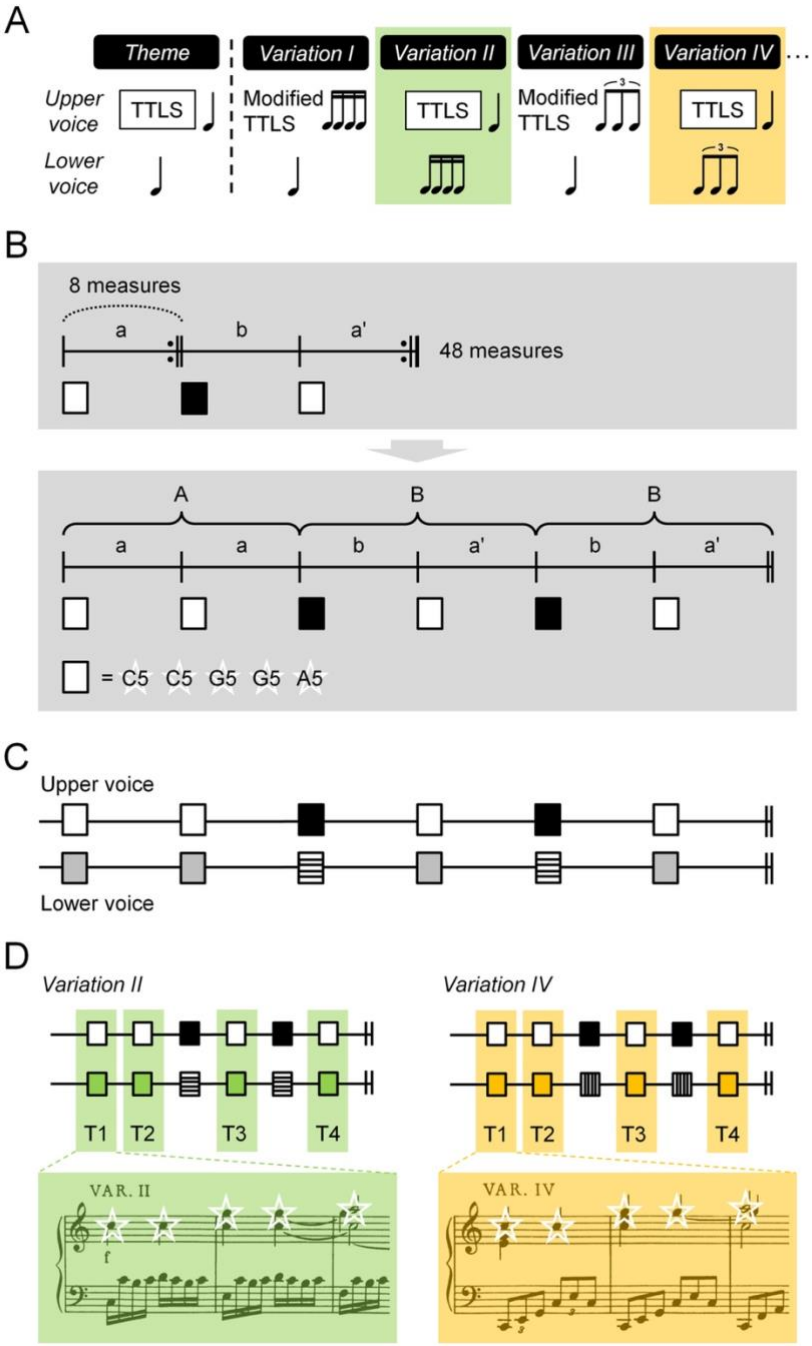

Fig. 2. Musical stimuli of Mozart's variation KV 265. A. *Variations I* to *IV* contain rhythmic variations on *Theme*. Compared to *Theme*, rhythmic patterns in the upper voices for *Variations I* and *III* are transformed while in *Variations II* and *IV* they are transformed into lower voices sharing the same TTLS melody. B. The top panel illustrates the structure of each variation involving repeat signs on the score. The bottom panel depicts the entire structure, comprising 48 measures of A (a + a) + B (b + a') + B (b + a'), as it is played. White squares stand for "C5-C5-G5-G5-A5" (marked by white stars in D), included in target phrases (T1, T2, T3, and T4 in D) repeated four times per variation,

while black squares denote the repeated cue phrases (Supplementary Fig. 1). C. Sequence of white and black squares in the upper voice is accompanied by lower voices. *Variations II* and *IV* both have two streams of voices. D. Different composition of notes in target phrases (T1 to T4) varies between the two variations, except for the upper voice denoted by the white squares. The target phrases are highlighted as green and orange shaded boxes, both contain the same white stars. White and green/orange squares indicate repetitions of time windows among the 48 measures of each variation in the upper and lower voices, respectively. Note: Scores adapted from *NMA Online: Neue Mozart-Ausgabe: Digitized Version* (https://dme.mozarteum.at/DME/nma/nmapub_srch.php?l=2). See Supplementary Fig. 1 for full scores, including T3 and T4.

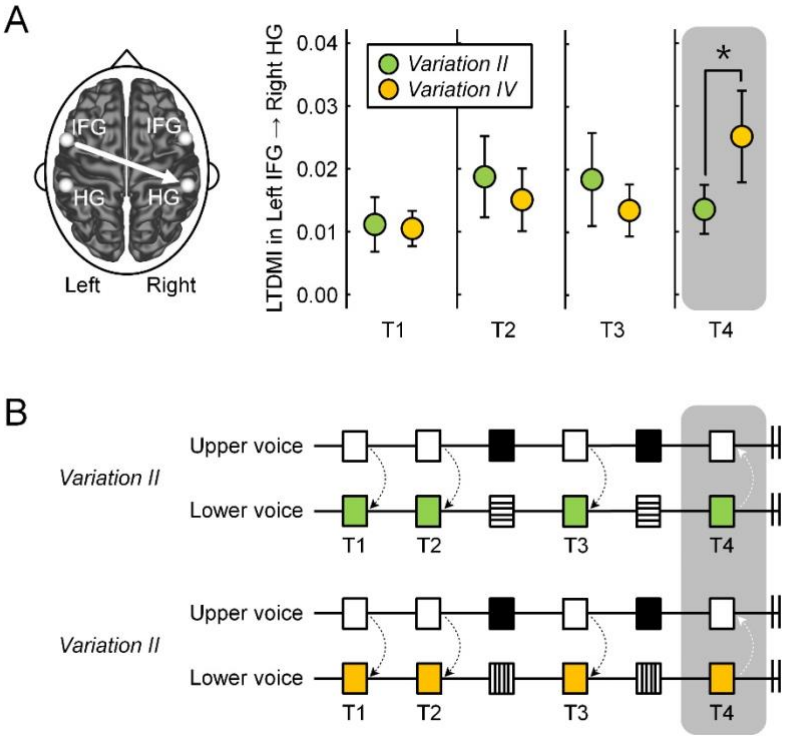

Fig. 3. Figure–ground reversal in *Variations II* and *IV* represented in changes in LTDMI values. A. Frontotemporal connectivity was significantly enhanced in T4 of *Variation IV* compared with *Variation II* (Wilcoxon signed-rank test, $Z = -2.112$, $P = 0.035$). T1 to T3 showed no significant differences. Error bar denotes the standard error mean (N = 25). *, $P < 0.05$. B. The figure–ground model illustrates the processing of upper and lower voices in *Variations II* and *IV* of Mozart's variation KV 265. The significant difference between *Variation II* and *Variation IV* indicates that the ground of the lower in T1 to T3 was reversed to the figure in T4. White squares denote phrases of the upper voice of the TTLS melody common to the two variations. Green and orange squares represent lower voices that differ between the variations. Black squares indicate cue phrases.