

Melodic sound enhances visual awareness of congruent musical notes, but only if you can read music

Minyoung Lee^a, Randolph Blake^{b,c,1}, Sujin Kim^a, and Chai-Youn Kim^{a,1}

^aDepartment of Psychology, Korea University, Seoul 136701, Korea; ^bDepartment of Psychological Sciences, Vanderbilt Vision Research Center, Vanderbilt University, Nashville, TN 37240; and ^cDepartment of Brain and Cognitive Sciences, Seoul National University, Seoul 151742, Korea

Contributed by Randolph Blake, May 14, 2015 (sent for review April 15, 2015; reviewed by David C. Burr and Sheng He)

Predictive influences of auditory information on resolution of visual competition were investigated using music, whose visual symbolic notation is familiar only to those with musical training. Results from two experiments using different experimental paradigms revealed that melodic congruence between what is seen and what is heard impacts perceptual dynamics during binocular rivalry. This bisensory interaction was observed only when the musical score was perceptually dominant, not when it was suppressed from awareness, and it was observed only in people who could read music. Results from two ancillary experiments showed that this effect of congruence cannot be explained by differential patterns of eye movements or by differential response sluggishness associated with congruent score/melody combinations. Taken together, these results demonstrate robust audiovisual interaction based on high-level, symbolic representations and its predictive influence on perceptual dynamics during binocular rivalry.

music | binocular rivalry | multisensory interaction | visual awareness | audio-visual congruence

Vision can be confusing, particularly when the retinal images inaugurating vision are ambiguous or underspecified with respect to their origins (1, 2). In some situations, confusion can be resolved by reliance on contextual information (3), prior experience (4), expectations (5), and/or implicit knowledge about structural regularities in our visual environment (6). In other situations, however, confusion resists resolution, in which case one can directly experience the dynamical nature of competition as conflicting perceptual interpretations vie for dominance (7). Among the visual phenomena exhibiting this behavior, none is more compelling than binocular rivalry (8).

When the two eyes view dissimilar monocular stimuli, stable binocular single vision gives way to interocular competition, characterized by unpredictable fluctuations in visual awareness of the rival stimuli (9, 10). Rivalry offers a way to study the inferential nature of perception when the brain is confronted with conflicting sensory evidence (11, 12). Within this context, recent studies imply that rivalry dynamics are governed by influences related to the likelihood of competing perceptual interpretations, such as motor actions (13), affective connotation (14–16), familiarity (17), and concomitant sensory input from other modalities (18–20). Here we examined the limits of susceptibility of rivalry to predictive influences by asking whether binocular rivalry dynamics depend on multisensory congruence between abstract representations familiar to individuals with expertise in that domain of abstraction. The domain we selected was music, for three reasons: (i) musical notation is symbolic; (ii) only people versed in music can decipher that notation; and (iii) melodic structure can be experienced via sight or sound.

Results

Experiment 1: Extended Tracking of Binocular Rivalry. Experiment 1 was implemented as a conventional binocular rivalry task in which participants pressed buttons to track alternating periods of dominance and suppression between dissimilar monocular displays viewed by the two eyes (Fig. 1A): One eye saw a musical score scrolling through a viewing window, and the other eye saw

a vertical, drifting grating. Participants tracked alternations in perception during 30-s trials administered in one of three audiovisual (AV) conditions: (i) listening to a melody that was congruent (CON) with the score, (ii) listening to a melody that was incongruent (INC), or (iii) listening to no sound (NS). The task itself required no familiarity with musical notation, allowing us to test people who can read music and those who cannot.

Our initial analyses focused on the relative predominance of the pair of rival stimuli (i.e., percentages of time the score or the grating was exclusively dominant). Because left eye and right eye predominance were equivalent, we pooled results across eyes for all statistical analyses of conditions and groups.

As summarized in Fig. 1B, musical scores in rivalry enjoyed significantly greater predominance than did grating among people able to read music [$t(22) = 3.71, P < 0.01$; Cohen's effect size index $d = 1.534$], whereas score and grating predominance were not different for nonreaders [$t(10) = 0.05, P = 0.956, d = 0.033$]. We next tailored our statistical analyses to examine predominance in more detail in only those persons who could read music. Paired t tests between the three AV conditions revealed that the predominance of scores was significantly greater in CON trials than in INC trials [$t(22) = 3.87, P < 0.001, d = 0.562$] or in NS trials [$t(22) = 0.78, P < 0.01, d = 0.602$].

For our purpose, it is essential to know whether these significant differences in score predominance for the AV conditions arose from lengthened durations of individual dominance of the scores and/or from abbreviated suppression of individual durations of the scores (i.e., dominance durations of the grating); either of those possibilities could promote the observed enhancement in

Significance

When left and right eyes disagree about what is being viewed, the brain resolves the disagreement by compromise: Visual awareness alternates between the two eyes' views over time. Called "binocular rivalry," these alternations in awareness are widely thought to reveal the usually implicit inferential nature of visual processing. In this study, we found that the perceptual dynamics defining rivalry are influenced by abstract, relational properties between visual and auditory sequences comprising musical melodies. However, this bisensory interaction impacts only the amount of time a visual sequence dominates awareness, with audiovisual interaction being powerless to hasten a visual sequence's emergence from invisibility into awareness. These findings situate the inferential processing putatively transpiring during rivalry prior to extraction of abstract, semantic information.

Author contributions: M.L., R.B., and C.-Y.K. designed research; M.L., R.B., S.K., and C.-Y.K. performed research; M.L. and S.K. analyzed data; and M.L., R.B., S.K., and C.-Y.K. wrote the paper.

Reviewers: D.C.B., University of Florence; and S.H., University of Minnesota.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

¹To whom correspondence may be addressed. Email: randolph.blake@vanderbilt.edu or chaikim@korea.ac.kr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1509529112/-DCSupplemental.

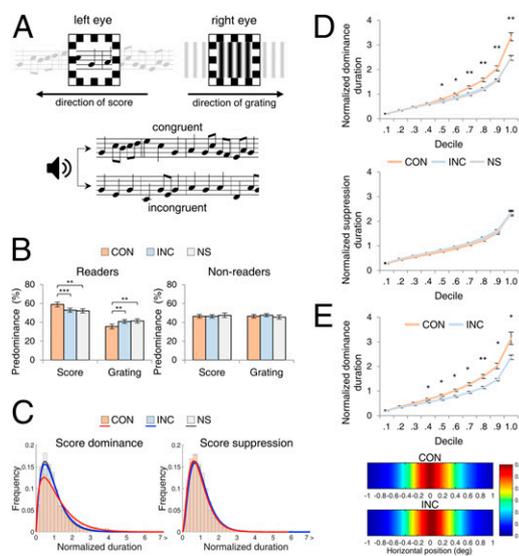


Fig. 1. Experiment 1. (A) Schematics of rival stimuli and AV conditions. (Upper) One eye viewed a single-note musical score scrolling right to left, and the other eye viewed a vertical grating scrolling left to right. (Lower) In some trials, the participant heard an auditory melody that either matched (CON) or did not match (INC) the visual score; in other trials (not illustrated) no melody was heard (NS). (B) Predominance plotted for readers and non-readers; pairs of predominance values for a given AV condition do not necessarily sum to 100%, because total trial duration included brief periods of incomplete dominance (i.e., mixtures, which comprised 5.6, 6.4, and 6.5% of the total viewing time for CON, INC, and NS conditions, respectively, values that are not significantly different when examined using ANOVA). Predominance values computed with mixtures excluded produce exactly the same pattern of results. $**P < 0.01$, $***P < 0.001$. Error bars denote SEM. (C) Frequency histograms showing the relative proportion of normalized dominance durations for score and for grating (i.e., score suppression) in readers in CON trials (orange), INC trials (blue), and NS trials (gray). (D) Decile plots for normalized score dominance durations and for normalized score suppression durations. $*P < 0.05$, $**P < 0.01$, FDR corrected. Error bars represent SEM. (E) Eye movement measurement in control experiment 1. (Upper) Decile plot of normalized score dominance durations for CON and INC conditions. $*P < 0.05$, $**P < 0.01$, FDR corrected. (Lower) Heat-maps showing horizontal gaze stability for CON and INC trials. The key indicates the mapping between color and incidence of fixation, with warm colors denoting greater incidence of fixation on the given spatial location of the rival target area.

predominance of musical scores. To compare the durations of dominance and suppression for these conditions, we first needed to take into account the substantial individual differences in overall durations of dominance existing among our participants: The average dominance durations spanned a threefold range (1.28–3.52 s), consistent with previous results (21, 22). To remove that source of variance from our analyses of dominance durations, we applied a widely used normalization procedure (23) whereby the individual dominance durations of scores for each participant were divided by the average dominance duration of scores and the individual suppression durations of scores were divided by the average suppression duration of scores. Frequency histograms of those normalized duration values (Fig. 1C) conform closely to the signature distribution characteristic of rivalry durations (24). Statistical analyses of these normalized durations yielded the same pattern of results as seen with predominance and, for that matter, with the actual dominance durations. Specifically, for the readers, normalized durations for scores seen in CON trials were significantly longer than those measured in INC trials [$t(22) = 4.04$, $P < 0.001$, $d = 1.454$] and those measured in NS trials [$t(22) = 3.96$, $P < 0.001$, $d = 1.524$]. Normalized score-suppressed durations did not differ significantly in these pairwise comparisons. We also analyzed normalized durations for the three AV conditions

within the nonreader group and found no statistically significant differences in score dominant or in score suppressed. The means and SEs of the raw and the normalized dominance and suppression durations for readers and nonreaders in each of the three AV conditions are provided in Table S1.

We next performed a more fine-grained analysis (25) of the normalized dominance and suppression durations measured in individual trials, the aim being to examine when in time those various AV conditions begin to differ from one another for the music readers. First, we sorted the individual normalized dominance durations and the individual normalized suppression durations measured during CON, INC, and NS trials from the shortest to the longest, separately for each participant. Then for each of those six conditions (two rivalry states \times three AV conditions), we divided the ordered sets of durations into deciles, i.e., 10 groups of durations, each with approximately equal numbers of ordered duration values. For each decile group we calculated the mean rivalry state durations and then averaged those individual decile means over observers to create the curves shown in Fig. 1D. Paired comparison t tests were performed on each of the CON/INC decile pairs, with P values corrected using the false-discovery rate (FDR) procedure applicable when performing multiple comparisons (26). The significant influence of CON did not emerge until dominance durations were equal to or longer than the dominance duration associated with the fifth decile. This finding makes sense, because congruence between a score and a melody requires experiencing multiple consecutive notes, a point we discuss further later. The CON and INC durations for the score-suppressed durations did not differ at any decile level.

Eye Movement Hypothesis. Both rival targets used in experiment 1 comprised contours moving in opposite directions, stimulus conditions that could have produced tracking eye movements (27), especially because we did not provide a fixation point (Methods). Could the differences in dominance durations measured in readers during CON and INC trials stem from differences in the pattern of eye movements evoked during those trials? That notion seemed unlikely, because the conditions of visual stimulation were identical for CON and for INC, but, to be sure, we replicated the conditions of experiment 1 in 20 new participants, all of whom were accomplished music readers, this time measuring eye movements throughout the course of each trial (see Methods and SI Results for details).

Considering first the psychophysical tracking results (Fig. 1E), dominance durations of the score were significantly longer in CON trials than in INC trials for all but the briefest decile categories, replicating the results from experiment 1.

Turning to the eye movement measures, those records disclose episodes of optokinetic reflex (OKR) movements correlated with the direction of motion associated with the reported direction of rival target motion, consistent with others' reports (28, 29). However, there also were periods devoid of OKR, in which the eyes were relatively stationary or executed small saccades during a rivalry state, making it impossible to predict reliably which stimulus was dominant (SI Results and Fig. S1). For our purpose, the mixed OKR results predicting the dominant stimulus was not a problem, because what we needed to know was whether the variability of horizontal eye position differed during CON trials and INC trials. In fact, we observed no discernible difference in the variability of eye positions in the two conditions (Fig. 1E and Fig. S2).

Experiment 2: Discrete Trial Binocular Rivalry. Congruence between auditory melody and visual musical score affected dominance durations for the score but had no significant influence on suppression durations for the score. Evidently, hearing music has no impact on the speed of recovery from suppression of a musical score, regardless whether music and score are CON or INC. This conclusion contrasts with other findings showing that contextual congruence (30, 31) and experience-based familiarity (32) can

influence how quickly a visual stimulus emerges from interocular suppression to achieve dominance. Hence, we wanted to examine more carefully the impact of INC and CON during phases of score suppression using a discrete trial probe technique that is arguably more sensitive than rivalry tracking when it comes to detecting weak, subliminal influences during rivalry (33, 34). This probe technique involves introducing CON or INC auditory melodies unpredictably when the musical score is dominant or suppressed (Fig. 2A).

Each trial began with dichoptic presentation of a nonmelodic score and a flickering radial grating, without accompanying sound. Readers and nonreaders of music reported rivalry state by pressing one of two assigned keys on the keyboard. In score-dominance trials, the nonmelodic score smoothly and rapidly morphed into one of seven possible melodic probe scores, with the change happening at the onset of the second dominance phase of the default score. In suppression trials, the transition from nonmelodic to probe score occurred at the onset of the second suppression phase of the nonmelodic score; because this transition occurred smoothly, participants were unaware of the change at the time of its occurrence. Trials in which transitions occurred at the onset of dominance or at the onset of suppression of the nonmelodic score were randomly intermixed, so observers could not anticipate when a transition would happen. Simultaneously with the transition onset, participants began hearing the melody corresponding to the newly introduced probe score (CON), a melody different from the newly introduced probe score (INC), or no melody at all (NS). A trial ended when the participant released the key signifying a change in perceptual dominance following introduction of the probe score. Thus each trial generated one phase duration, which could be either a dominance duration for the probe score or a suppression duration for the probe score.

Fig. 2B presents frequency histograms for all normalized dominance and suppression durations measured in the probe phase of each trial. Rival state durations measured using single trial probes, as an aggregate, conform to the characteristic gamma-like distribution typical of durations measured using extended rivalry tracking (24). Fig. 2C shows decile plots for CON, INC, and NS conditions, for probes presented during dominance and during suppression for readers. As found in experiment 1, CON trials produced significantly longer dominance durations than did INC trials for decile categories at or beyond the median value, again confirming that CON boosts dominance of the musical score. Of primary interest for our purposes, suppression durations of the probe scores were not significantly different within any of the decile groups. As in experiment 1, nonreaders did not exhibit significant differences across the three AV conditions.

Having again failed to find an effect of congruency on suppression durations, we wondered whether some other factor might prevent us from detecting a weak but genuine effect of score congruency on suppression durations. As noted earlier, congruency of auditory melody and visual score can be discerned only when there are multiple successive tones and visual notes available to be matched. Perhaps, we reasoned, a genuine effect of congruency during suppression is being obscured by inclusion of very short-state durations that cannot possibly contribute to a congruency effect. To examine that possibility, we reanalyzed the dominance and suppression durations for CON and INC conditions based on the number of notes actually presented in each trial. The pair of histograms in Fig. 2D plots the number of trials in which at least n notes were visually and auditorily presented before a state transition. Plotted in this way, the data can be construed as the outcome of a series of Bernoulli trials, with the resulting geometric distribution plotted as a function of duration, except that the unit of time in Fig. 2D is the number of notes presented. The relevant comparison is between the CON and INC trials. Looking first at the dominance trials (i.e., probe score and accompanying sound presented when the probe score was dominant), the differences between CON and INC dominance

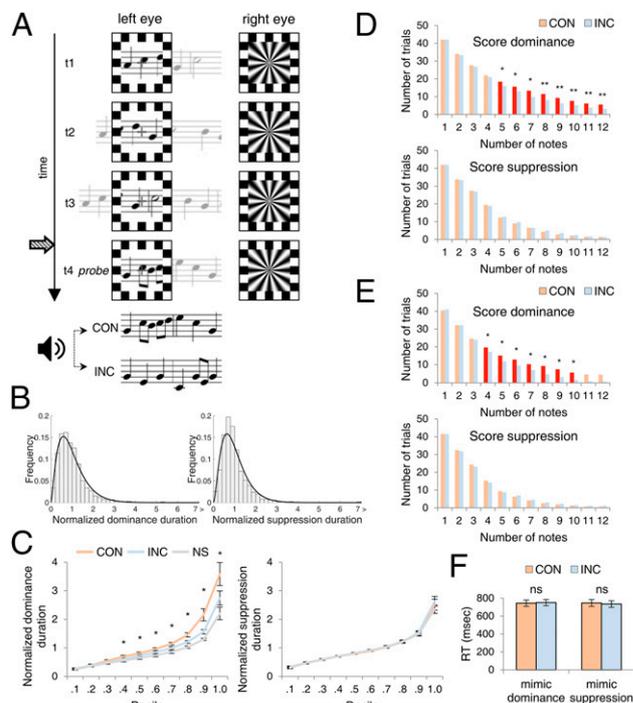


Fig. 2. Experiment 2. (A) Schematics of rival stimuli (t_1 , t_2 , and t_3 denote time) and AV conditions. One eye viewed a musical score drifting leftward while the other eye viewed a counterphase flickering radial grating. Following an unpredictable number of state changes (signified by participant's button presses), the initially random score smoothly morphed into a melodic probe score (signified by the time label t_4). Concurrently with probe presentation, the participant began hearing an auditory melody that either matched (CON) or did not match (INC) the probe score. In other trials, probe presentations were unaccompanied by auditory melody (NS). (B) Frequency histograms of the relative proportion of the normalized score dominance durations and score suppression durations in both readers and nonreaders, collapsed over AV conditions. (C) Decile plots for the normalized score dominance durations and normalized score suppression durations. $*P < 0.05$, FDR corrected. Error bars denote SEM. (D) Frequency histograms of trials in experiment 2 categorized by the number of notes presented visually and auditorily for score dominance and for score suppression conditions in readers. Bolded bars and symbols denote categories in which CON and INC were significantly different in terms of incidence of number of trials. $*P < 0.05$, $**P < 0.01$ FDR corrected. (E) Results in same format as D, obtained from a new group of participants. $*P < 0.05$, FDR corrected. (F) Reaction times (RTs) to probe scores presented in rivalry mimic trials in which the probe score followed appearance of the nonsense melody score (mimicking score dominance) or followed the appearance of the radial grating (mimicking score suppressed). Response times did not differ between the two AV conditions (ns, not significant). Error bars indicate SEM.

durations are statistically significant for trials in which anywhere from 5 to 12 musical notes had been seen and heard; before that, CON and INC trials were statistically indiscriminable in duration. Evidently the realization of the congruency or lack of congruency between auditory and visual melodic structure emerges only after some critical amount of time, on the order of a couple of seconds. In other words, CON trials tended to last longer when observers had the opportunity to hear and to see five or more notes. Looking next at the suppression trials (i.e., probe score and accompanying sound presented immediately following suppression of the initial musical score), one sees an equivalence of CON and INC trials at all note values. So, considering only at trials in which readers experienced multiple musical notes does not reveal a subtle reduction in suppression durations associated with melodic congruency.

Finally, we tried excluding state durations that were too brief to ensure the presentation of at least two visual and two auditory

notes; those minimum durations were defined individually for each of the seven musical sequences used in the experiment. This exclusion rule was imposed on the entire set of actual rival state durations—dominance and suppression—following the introduction of the auditory melody, regardless whether they were CON or INC trials. Excluded trials averaged 20% of the total across all participants. The removal of very brief durations had no impact on the robust difference between CON and INC dominance durations and, moreover, did not uncover a weak suppression effect. Summary values of those state durations, both unpruned and pruned, are presented in [Table S2](#).

Sluggish Response Hypothesis. Given the failure to find any trace of auditory modulation of suppression durations, we felt compelled to consider whether the effect of congruence on dominance durations might, in fact, be attributable to response sluggishness and not to prolonged perceptual dominance. According to this hypothesis, readers are slower to recognize and respond to unpredictable transitions from score to grating in CON trials than in INC trials, an effect that would not be seen in trials involving the transition from grating to score. Perhaps, for example, congruent audiovisual stimulation is more engaging, leading to slower button releases and thus to seemingly longer dominance durations. To evaluate this hypothesis, we tested 13 music readers (11 new to the task) using exactly the same discrete trial procedure used in experiment 2. The results for these CON and INC trials replicated the pattern found in the original group of participants (Fig. 2E). Using those measured dominance and suppression durations, we then created rivalry mimic trials for each participant that entailed actually presenting the score and the grating alternately to the separate eyes in a sequence matching that reported by that participant during the actual rivalry procedure (*SI Methods* and Fig. S3). As before, participants tracked the alternating appearance of the two rival stimuli by pressing one of two buttons. Means and SDs of the response latencies associated with the final transition of a sequence were computed for CON and INC trials associated with the ersatz dominance condition (real transition from score to grating) and the ersatz suppression condition (real transition from grating to score). For both the mimicked dominance and the mimicked suppression trials, reaction times in the CON and INC conditions were indistinguishable (Fig. 2F), implying that reaction times per se are not responsible for the difference in dominance durations between the two conditions measured during genuine rivalry.

Discussion

Melodic congruence between what is seen and what is heard impacts perceptual dynamics during binocular rivalry in people who can read music. However, this interaction between seeing and hearing was not observed when the musical score was suppressed from awareness, and under no conditions was it seen in people without expertise in music. This effect of congruence in music readers cannot be attributed to differential patterns of eye movements or to response sluggishness associated with CON score/melody combinations.

Our results add to the body of evidence showing that visual awareness during interocular competition is influenced by non-visual information conveyed by other sensory modalities (35–41). Our results go further, however, by demonstrating robust AV interaction based on high-level, symbolic representations, i.e., musical notation and melodic structure. It is not surprising that this robust interaction is enjoyed only by individuals with expertise acquired through musical training, in contrast, for example, to the performance enhancement exhibited by virtually all people when listening to speech sounds while looking at lip movements (35, 39).

Is it far-fetched to interpret the current findings as resulting partly from preexisting perceptual characteristics of individuals trained in music? It has been suggested that musicians exhibit distinct auditory (42, 43), verbal (44), and broad cognitive (45)

abilities and/or systems. In particular, musical training, whether singing or playing an instrument, requires developing perception–action coordination, which may promote unusually strong connections between perception and action systems in the brain (46, 47). Storage and retrieval of musical melodies may be subserved by auditory-motor pathways including the premotor area (48). In this vein, our results could arise in part from feedback from implicit motor activity evoked in musicians while viewing musical notes (49). From other work we know that what one is doing can impact dominance in binocular rivalry (13) and that actions relevant to what is perceived influence perceptual dynamics during bistable perception (50); perhaps this influence is at play in our situation.

The multisensory interaction demonstrated in our study almost certainly does not arise from sensory-neural convergence of auditory and visual signals, i.e., the kind of convergence that has been well-established in single-cell sensory neurophysiology (51–53) and in human brain imaging (54). Instead, the interactions implicated in our study more likely entail information combination of the sort discussed in the context of cue combination across modalities (55, 56). In those situations the process of combination is construed as a form of probabilistic inference with dynamic weightings of different sources of information being governed by their reliability and likelihood. To interpret our results in that framework, however, would mean that likelihood, as signified by CON vs. INC, can be established only during dominance phases of rivalry, not during suppression.

Alternatively, one could imagine that congruent melodic sound and visual score more strongly engage the attention of a musically trained individual, resulting in a strengthened neural representation for that combination (9). After all, attention can modulate the visual awareness of a stimulus engaged in binocular competition (57–59), and this modulation could be a mediating factor in our study. A weakness of this hypothesis, however, is the equally plausible expectation that incongruence should attract attention and therefore should more strongly boost dominance of mismatched melody/score combinations. Indeed, in other situations incongruity does indeed promote dominance of a rival stimulus (60).

The AV congruence effect discovered in our work may arise from neural events within high-level, amodal brain areas such as the posterior portion of the superior temporal sulcus (61) and the prefrontal cortex (62). These brain regions also are implicated in music learning (63). Activity in these high-level regions, in turn, might have feedback influence on early visual cortex (64). Complex auditory information and even imagery can be decoded from activity in the primary visual cortex in the absence of retinal input (65). Hence, an interplay between the high-level multisensory and the early visual areas could form part of the neural network promoting visual predominance of musical notation accompanied by congruent melodic sound.

Nonetheless, we are left with an unanswered question: Why does hearing a melody have no impact on suppression durations associated with viewing a musical score, regardless whether melody and score are congruent or incongruent? After all, in other circumstances suppression durations of a visual stimulus are impacted by the properties of that stimulus, including its physical characteristics (9), its configural properties (30, 66), its affective connotation (67, 68), its lexical familiarity (32), and its social connotation (69). However, in all these instances the features defining the suppressed stimulus are plausibly being registered and remain available for analysis (70), albeit with reduced fidelity or signal strength (34). There is no reason why this should not be true for the visual features—contour size, orientation and contrast polarity—comprising notes contained in the musical scores used as a rival stimulus in our experiments. However, melody is embodied in the unfolding temporal structure of successive notes; in this sense, melody is an emergent property of a sequence of notes. Abstracting the lexical semantics of music is comparable in processing complexity to the abstraction of meaning from a sequence of words comprising a

sentence. Evidently, without conscious access to the relational patterns of notes portrayed visually in a score, the emergent property of melody cannot be realized and therefore cannot be combined with auditory melody during suppression phases of the score; bisensory melodic integration is limited to periods of dominance of the score. During suppression, in other words, the visually defined semantic content within a musical score is literally out of sight and out of mind.

Methods

General Methods.

Participants. Different portions of this study were conducted at Korea University, at Seoul National University, and at Vanderbilt University. All volunteer participants had normal or corrected-to-normal visual acuity and good stereopsis; participants were not screened based on eye dominance because rival target/eye configuration was counterbalanced over trials. Each person gave written informed consent on forms approved by the Institutional Review Board (IRB) of Korea University (1040548-KU-IRB-13-149-A-2) or Vanderbilt University (IRB 040915).

Apparatus and stimuli. Stimuli were created using Matlab in concert with the Psychophysics Toolbox-3 (71, 72). At Korea University and Seoul National University, visual stimuli were presented on a 19-inch CRT monitor (1,024 × 768 resolution, 60-Hz frame rate, 53 cm viewing distance at Korea University, 60 cm in Seoul National University), and at Vanderbilt University, visual stimuli were presented on a 20-inch CRT monitor (1,024 × 768 resolution, 100-Hz frame rate, 80.5 cm viewing distance). In all setups, participants viewed the monitor through a mirror stereoscope with the head stabilized by a head/chin-rest. Sound was delivered over headphones.

Experiment 1.

Participants. Twenty-three music readers [15 females, age 22.3 ± 2.5 (SD) y] and 11 nonreaders [seven females, age 26.1 ± 9.1 (SD) y] participated in the experiment, some at Korea University and others at Vanderbilt University.

Apparatus and stimuli. Dichoptically viewed rival targets were generated on the two halves of the video monitor, so that one eye viewed one of seven possible musical scores and the other eye viewed a 2.1 cycle per degree, 60%-contrast, vertical, drifting sinusoidal grating. Auditory stimuli comprised seven melodies, each corresponding to one of the seven musical scores. Scores (dark gray notes against a white background, ~60% contrast) were excerpted from vocal scores (monophonic), with notes edited in size, note density, and rhythm as needed to ensure visual clarity. The visual scores and the auditory sound files corresponding exactly to those scores were generated with the music notation software Sibelius (Avid Technology, Inc.). In each trial one of the seven scores and a vertical grating were foveally viewed within a pair of dichoptic windows framed by a checkerboard texture that served to promote stable binocular alignment (Fig. 1A). The inner and outer borders of the windows subtended a visual angle of $1.9^\circ \times 1.9^\circ$ and $2.66^\circ \times 2.66^\circ$, respectively. The score scrolled from right to left within its window at $\sim 1.2^\circ/s$, and the grating drifted within its window at approximately the same speed in the opposite direction.

Design and procedure. While dichoptically viewing the rival stimuli, participants used keys to track successive periods of exclusive dominance of the score and of the grating, pressing both keys when mixtures of the score and the grating were experienced. Trial durations varied from 30 to 42 s depending on which of the seven musical sequences was used in that trial, and each trial comprised one of three possible AV conditions of equal proportion. In CON trials, participants heard the auditory melody synchronized with the streaming visual score; in INC trials they heard an auditory melody different from the visual score (drawn from the pool of six remaining melodies); and in NS trials no melody was heard. Each visual score was repeated six times and presented an equal number of times to each eye, yielding a total of 42 trials per observer. For each trial we discarded the last rivalry state duration if the observer was pressing a key at the time the trial ended, to avoid including truncated durations in the data analysis. A fixation mark was not included, but all individuals were instructed to maintain their gaze within the display window throughout the trial. They were told to expect musical sound to be heard in some trials but not others. Before formal data collection, participants were given practice tracking rivalry between score and grating in the absence of sound.

Control Experiment 1: Eye Movements.

Participants. Fourteen volunteers were tested at Korea University, and six were tested at Seoul National University [17 females, age 21.8 ± 2.1 (SD) y]. Given the purpose of this experiment (see text), we tested only people who could

read music. Some of those individuals were tested during the course of a related study comparing rivalry dynamics for melody/score combinations in musicians with absolute pitch and in musicians without absolute pitch.

Apparatus, stimuli, and procedure. Apparatus, stimuli, and procedure were identical to those in experiment 1, with the addition of recording eye movement throughout each trial. Those records were made using a video eye-tracking system that sampled the left eye's position (1,000 Hz) using the centroid mode [EyeLink 1000, SR Research, controlled by the EyeLink toolbox for Matlab (73)]. Initial calibration and two subsequent recalibrations during the test session were performed using the conventional five-point calibration routine that forms part of the EyeLink software package. Participants viewed the screen through a mirror stereoscope with their heads stabilized by a head/chin-rest. As before, there was no fixation point, but participants were instructed to look at the display window throughout the trial. For the analysis of eye-movement data, we excluded periods of eye blinks as well as records measured between 100 ms before and 100 ms after each eye blink (those periods comprised 10% of the total trial duration). A total of 42 trials were broken into blocks of 14 trials each, and during a block of trials observers were not allowed to remove their heads from the head/chin-rest. After each block, observers were given a short rest period followed by eye tracker recalibration before beginning the next block.

Experiment 2.

Participants. Fifteen music readers (10 females, age 22.1 ± 3.2 (SD) y) and 15 nonreaders [seven females, age 26.2 ± 9.8 (SD) y] who had not participated in experiment 1 were recruited at Vanderbilt University. In this experiment we validated participants' self-reported ability/inability to read music by giving each individual a music quiz consisting of multiple-choice questions, each with four possible answer choices that involved identification of a specific, visually presented musical note.

Apparatus and stimuli. Stimuli and apparatus were identical to those used in experiment 1 except that (i) the rival grating was a 100% contrast, 5-Hz counterphase flickering radial grating (Fig. 2A), selected because it had no tendency to induce lateral eye movements; (ii) each trial started with one eye viewing a nonmelodic musical score consisting of a quasi-random sequence of notes that transitioned into one of seven actual musical scores (the probe score) after several rivalry alternations (Fig. 2A); (iii) a fixation point in the center of the window displaying the musical score, a location coinciding with the point within the other window where the radial grating came to a point; and (iv) musical scores the notes of which were convolved with a Gaussian filter ($\sigma = 0.08^\circ$ visual angle) to reduce the sharpness of their edges without compromising legibility.

Each participant completed three blocks comprising 84 discrete rivalry trials. At the beginning of a trial, the nonmelodic score and the flickering radial grating were presented dichoptically without accompanying sound. Each rival target was presented an equal number of times to the two eyes. Equal numbers of the three AV conditions were intermixed randomly within a block of 42 trials, and each block was repeated three times. A trial ended as soon as the participant indicated a change in perceptual dominance following introduction of the probe score; thus each trial generated one phase duration, which could be either a dominance duration for the probe score or a suppression duration for the probe score (i.e., those trials in which dominance of the radial grating was being reported). The end of each trial was followed by an enforced 1-s interval before the observer could tap a key to initiate the next trial.

Control Experiment 2: Response Sluggishness.

Participants. Thirteen Korea University students [nine females, age 23.5 ± 2.2 (SD) y] participated in this control experiment, including two who also had participated in experiment 1. We purposefully tested only music readers, because this experiment focused on the congruence effect found in readers only.

Apparatus and stimuli. Stimuli and apparatus were identical to those in experiment 2, with the addition of yoked rivalry trials in which rivalry alternations were mimicked by physically alternating between the radial grating and the score (SI Methods and Fig. S3).

ACKNOWLEDGMENTS. We thank Gordon Logan and Isabel Gauthier for helpful comments on this project. This work was supported by Grants NRF-2013R1A1A1010923 and NRF-2013R1A2A2A03017022 from the Basic Science Research Program through the National Research Foundation of Korea (NRF) [funded by the Ministry of Science, Information and Communications Technologies (ICT) and Future Planning] and by National Institutes of Health Core Grant P30-EY008126.

- Gregory RL (1980) Perceptions as hypotheses. *Philos Trans R Soc Lond B Biol Sci* 290(1038):181–197.
- Knill DC, Richards W (1996) *Perception as Bayesian Inference* (Cambridge Univ Press, Cambridge, UK).
- Albright TD, Stoner GR (2002) Contextual influences on visual processing. *Annu Rev Neurosci* 25:339–379.
- Bar M (2004) Visual objects in context. *Nat Rev Neurosci* 5(8):617–629.
- Sterzer P, Frith C, Petrovic P (2008) Believing is seeing: Expectations alter visual awareness. *Curr Biol* 18(16):R697–R698.
- Marr D (1970) A theory for cerebral neocortex. *Proc R Soc Lond B Biol Sci* 176(1043):161–234.
- Blake R, Logothetis N (2002) Visual competition. *Nat Rev Neurosci* 3(1):13–21.
- Fox R, Blake R (2005) Psychophysical magic: Rendering the visible ‘invisible’. *Trends Cogn Sci* 9(8):381–388.
- Levelt WJM (1965) *On Binocular Rivalry* (Royal Van Gorcum, Assen, The Netherlands).
- Fox R, Herrmann J (1967) Stochastic properties of binocular rivalry alternations. *Percept Psychophys* 2(9):432–436.
- Hohwy J, Roepstorff A, Friston K (2008) Predictive coding explains binocular rivalry: An epistemological review. *Cognition* 108(3):687–701.
- Leopold DA, Logothetis NK (1999) Multistable phenomena: Changing views in perception. *Trends Cogn Sci* 3(7):254–264.
- Maruya K, Yang E, Blake R (2007) Voluntary action influences visual competition. *Psychol Sci* 18(12):1090–1098.
- Alpers GW, Pauli P (2006) Emotional pictures predominate in binocular rivalry. *Cogn Emot* 20(5):596–607.
- Yoon KL, Hong SW, Joormann J, Kang P (2009) Perception of facial expressions of emotion during binocular rivalry. *Emotion* 9(2):172–182.
- Wilbertz G, van Slooten J, Sterzer P (2014) Reinforcement of perceptual inference: Reward and punishment alter conscious visual perception during binocular rivalry. *Front Psychol* 5:1377.
- Yu K, Blake R (1992) Do recognizable figures enjoy an advantage in binocular rivalry? *J Exp Psychol Hum Percept Perform* 18(4):1158–1173.
- Conrad V, Bartels A, Kleiner M, Noppeney U (2010) Audiovisual interactions in binocular rivalry. *J Vis* 10(10):1–15.
- Lunghi C, Alais D (2013) Touch interacts with vision during binocular rivalry with a tight orientation tuning. *PLoS ONE* 8(3):e58754.
- Zhou W, Jiang Y, He S, Chen D (2010) Olfaction modulates visual perception in binocular rivalry. *Curr Biol* 20(15):1356–1358.
- Carter OL, Pettigrew JD (2003) A common oscillator for perceptual rivalries? *Perception* 32(3):295–305.
- Hancock S, Gareze L, Findlay JM, Andrews TJ (2012) Temporal patterns of saccadic eye movements predict individual variation in alternation rate during binocular rivalry. *Iperception* 3(1):88–96.
- Logothetis NK, Leopold DA, Sheinberg DL (1996) What is rivaling during binocular rivalry? *Nature* 380(6575):621–624.
- Brascamp JW, van Ee R, Pestman WR, van den Berg AV (2005) Distributions of alternation rates in various forms of bistable perception. *J Vis* 5(4):287–298.
- Ratcliff R (1979) Group reaction time distributions and an analysis of distribution statistics. *Psychol Bull* 86(3):446–461.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1):289–300.
- Valsecchi M, Gegenfurtner KR, Schütz AC (2013) Saccadic and smooth-pursuit eye movements during reading of drifting texts. *J Vis* 13(10):1–20.
- Fox R, Todd S, Bettinger LA (1975) Optokinetic nystagmus as an objective indicator of binocular rivalry. *Vision Res* 15(7):849–853.
- Hayashi R, Tanifuji M (2012) Which image is in awareness during binocular rivalry? Reading perceptual status from eye movements. *J Vision* 12(3):5, 1–11.
- Wang L, Weng X, He S (2012) Perceptual grouping without awareness: Superiority of Kanizsa triangle in breaking interocular suppression. *PLoS ONE* 7(6):e40106.
- Mudrik L, Breska A, Lamy D, Deouell LY (2011) Integration without awareness: Expanding the limits of unconscious processing. *Psychol Sci* 22(6):764–770.
- Jiang Y, Costello P, He S (2007) Processing of invisible stimuli: Advantage of upright faces and recognizable words in overcoming interocular suppression. *Psychol Sci* 18(4):349–355.
- Lin Z, He S (2009) Seeing the invisible: The scope and limits of unconscious processing in binocular rivalry. *Prog Neurobiol* 87(4):195–211.
- Ling S, Blake R (2009) Suppression during binocular rivalry broadens orientation tuning. *Psychol Sci* 20(11):1348–1355.
- Alsius A, Munhall KG (2013) Detection of audiovisual speech correspondences without visual awareness. *Psychol Sci* 24(4):423–431.
- Chen Y-C, Yeh S-L, Spence C (2011) Crossmodal constraints on human perceptual awareness: Auditory semantic modulation of binocular rivalry. *Front Psychol* 2:212, 10.3389/fpsyg.2011.00212.
- Kang MS, Blake R (2005) Perceptual synergy between seeing and hearing revealed during binocular rivalry. *Psychologia* 32:7–15.
- Lunghi C, Morrone MC, Alais D (2014) Auditory and tactile signals combine to influence vision during binocular rivalry. *J Neurosci* 34(3):784–792.
- Plass J, Guzman-Martinez E, Ortega L, Grabowecy M, Suzuki S (2014) Lip reading without awareness. *Psychol Sci* 25(9):1835–1837.
- van Ee R, van Boxtel JJ, Parker AL, Alais D (2009) Multisensory congruency as a mechanism for attentional control over perceptual selection. *J Neurosci* 29(37):11641–11649.
- Lunghi C, Binda P, Morrone MC (2010) Touch disambiguates rivalrous perception at early stages of visual analysis. *Curr Biol* 20(4):R143–R144.
- Hyde KL, et al. (2009) Musical training shapes structural brain development. *J Neurosci* 29(10):3019–3025.
- Schlaug G, et al. (2009) Training-induced neuroplasticity in young children. *Ann N Y Acad Sci* 1169(1):205–208.
- Franklin MS, et al. (2008) The effects of musical training on verbal memory. *Psychol Music* 36(3):353–365.
- Jacoby N, Ahissar M (2013) What does it take to show that a cognitive training procedure is useful? A critical evaluation. *Prog Brain Res* 207:121–140.
- Novembre G, Keller PE (2014) A conceptual review on action-perception coupling in the musicians’ brain: What is it good for? *Front Hum Neurosci* 8:603, 10.3389/fnhum.2014.00603.
- Zatorre RJ, Chen JL, Penhune VB (2007) When the brain plays music: Auditory-motor interactions in music perception and production. *Nat Rev Neurosci* 8(7):547–558.
- Rauschecker JP (2014) Is there a tape recorder in your head? How the brain stores and retrieves musical melodies. *Front Syst Neurosci* 8:149, 10.3389/fnsys.2014.00149.
- Reichenbach A, Franklin DW, Zatzka-Haas P, Diedrichsen J (2014) A dedicated binding mechanism for the visual control of movement. *Curr Biol* 24(7):780–785.
- Beets IAM, et al. (2010) Online action-to-perception transfer: Only percept-dependent action affects perception. *Vision Res* 50(24):2633–2641.
- Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. *Science* 221(4608):389–391.
- Maier JX, Neuhoff JG, Logothetis NK, Ghazanfar AA (2004) Multisensory integration of looming signals by rhesus monkeys. *Neuron* 43(2):177–181.
- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci* 25(20):5004–5012.
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10(11):649–657.
- Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415(6870):429–433.
- Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 14(3):257–262.
- Dieter KC, Tadin D (2011) Understanding attentional modulation of binocular rivalry: A framework based on biased competition. *Front Hum Neurosci* 5(155):1–12.
- Paffen CLE, Alais D (2011) Attentional modulation of binocular rivalry. *Front Hum Neurosci* 5:105, 10.3389/fnhum.2011.00105.
- Zhang P, Jiang Y, He S (2012) Voluntary attention modulates processing of eye-specific visual information. *Psychol Sci* 23(3):254–260.
- Mudrik L, Deouell LY, Lamy D (2011) Scene congruency biases binocular rivalry. *Conscious Cogn* 20(3):756–767.
- Naumer MJ, et al. (2011) Investigating human audio-visual object perception with a combination of hypothesis-generating and hypothesis-testing fMRI analysis tools. *Exp Brain Res* 213(2-3):309–320.
- Diehl MM, Romanski LM (2014) Responses of prefrontal multisensory neurons to mismatching faces and vocalizations. *J Neurosci* 34(34):11233–11243.
- Zimmerman E, Lahav A (2012) The multisensory brain and its ability to learn music. *Ann N Y Acad Sci* 1252(1):179–184.
- Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on ‘sensory-specific’ brain regions, neural responses, and judgments. *Neuron* 57(1):11–23.
- Vetter P, Smith FW, Muckli L (2014) Decoding sound and imagery content in early visual cortex. *Curr Biol* 24(11):1256–1262.
- Zhou G, Zhang L, Liu J, Yang J, Qu Z (2010) Specificity of face processing without awareness. *Conscious Cogn* 19(1):408–412.
- Jiang Y, Costello P, Fang F, Huang M, He S (2006) A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proc Natl Acad Sci USA* 103(45):17048–17052.
- Yang E, Zald DH, Blake R (2007) Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion* 7(4):882–886.
- Stein T, Senju A, Peelen MV, Sterzer P (2011) Eye contact facilitates awareness of faces during interocular suppression. *Cognition* 119(2):307–311.
- Blake R, Tadin D, Sobel KV, Raissian TA, Chong SC (2006) Strength of early visual adaptation depends on visual awareness. *Proc Natl Acad Sci USA* 103(12):4783–4788.
- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10(4):433–436.
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat Vis* 10(4):437–442.
- Cornelissen FW, Peters EM, Palmer J (2002) The EyeLink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behav Res Methods Instrum Comput* 34(4):613–617.