

## WHAT DRIVES NARRATIVE ENGAGEMENT WITH MUSIC?

---

J. DEVIN MCAULEY  
Michigan State University

PATRICK C. M. WONG  
Chinese University of Hong Kong

LUCAS BELLAICHE  
University of Arkansas

ELIZABETH HELLMUTH MARGULIS  
Princeton University

**ALTHOUGH PEOPLE ACROSS MULTIPLE CULTURES** have been shown to experience music narratively, it has proven difficult to disentangle whether narrative dimensions of music derive from learned extramusical associations within a culture or from less experience-dependent elements of the music, such as musical contrast. Toward this end, two experiments investigated factors contributing to listeners' narrative engagement with music, comparing the narrative experiences of Western and Chinese instrumental music for listeners in two suburban locations in the United States with those of listeners living in a remote rural village in China with different patterns of musical exposure. Supporting an enculturation perspective where learned extramusical associations (i.e., Topicality) play an important role in narrative perceptions of music, results from the first experiment show that for Western listeners, greater Topicality, rather than greater Contrast, increases narrative engagement, as long as listeners have sufficient exposure to its patterns of use within a culture. Strengthening this interpretation, results from the second experiment, which directly manipulated Topicality and Contrast, show that reducing an excerpt's Topicality, but not its Contrast reduces listeners' narrative engagement.

Received: September 4, 2020, accepted March 21, 2021.

**Key words:** music, narrative, imagination, culture, communication

---

**T**HERE ARE MANY DIFFERENT WAYS TO EXPERIENCE music (see Tuuri & Eerola, 2012, for a proposed taxonomy, and Feld, 1984, for a sensitive

account of the relevant dimensions). A person might listen to a melody and respond kinesthetically by tapping their foot, or they might appreciate its beauty, or even, recognizing the melody as their ringtone, run to pick up their phone. Listening to the same melody, they might instead imagine a character or setting within an imagined unfolding story—the melody might sound like a cat chasing a mouse or a princess approaching the moment of coronation. This last type of listening, one of multiple possible ways of interacting with music, is termed narrative engagement with music, and can be measured by the Narrative Engagement Scale, a quick four-question inventory assessing the story's vividness, immediacy, clarity, and ease of imagining that has been developed and validated at multiple research sites (Margulis, Wong, Simchy-Gross, & McAuley, 2019).

Narrative imaginings, together with visual imagery, fall into a broader class of listening experiences that Herbert (2011) terms *musical daydreams*—experiences of music that are “marked by a fluctuating distributed attentional focus.” These kinds of experiences do not constitute arbitrary mind wandering, but rather stimulus- and context-driven episodes in which the sounds seem to shape multimodal trajectories through an imagined space. Instead of hearing a cello passage alternate with a trombone *qua* an instrumental exchange, a narratively engaged listener might hear it as a thief struggling to evade the authorities. The sounds are there, but they point to happenings in an imagined story rather than to themselves. Music's capacity to gesture toward external events is familiar from film music (Kassabian, 2001; Tan, Cohen, Lipscomb, & Kendall, 2013.) Yet we also know that music commonly triggers imagery in the absence of any visual component (Küssner & Eerola, 2019); that such imaginings can generate high levels of absorption (Presicce & Bailes, 2019; Vroegh, 2019), and that there can be substantial intersubjectivity in imaginings sustained in response to particular excerpts (Huovinen & Kaila, 2015; Margulis et al., 2019; McAuley, Wong, Mamidipaka, Phillips, & Margulis, 2021; Tagg & Clarida, 2003).

Previous work has shown that people at multiple research sites in the US and China can engage narratively with previously unheard examples of instrumental music, and that individual excerpts generate shared, concrete stories within but not across cultures (Margulis et al., 2019; McAuley et al., 2021). Individual

excerpts, in other words, tend to generate specific stories that people within a culture broadly recognize as appropriate—one excerpt might inspire stories of an animal chase, for example, and another stories of grief and loss. While previous research has established that narrative engagement occurs, no empirical research has yet explored the factors that contribute to narrative perceptions of music. What is happening in the music that differentiates excerpts to which participants respond narratively from excerpts to which they do not?

Music theory offers tools for thinking about this question. One account sees narrative engagement as a fundamental human capacity driven by specific stimulus features independent of listener exposure and enculturation. Proponents of this account see Heider and Simmel's (1944) studies as evidence of a tendency to understand abstract stimuli narratively under particular circumstances. Participants in these studies understood moving shapes as intentional agents interacting in a dramatic story, especially when changes in the movements of the shapes seemed to violate intuitive rules of physics (e.g., fail to conform to principles of inertia or gravity). Similarly, writers on music have suggested that people resort to narrative interpretations of musical sounds when those sounds violate expectations in some way; for example, by introducing substantial contrast with what came before (loud after soft, for example, or slow after fast, or a solo flute following an orchestral tutti; Almén, 2008). According to this account, narrative engagement should increase with the amount of *contrast* present in the music, regardless of the background and previous experiences of the listener.

Another account views associational pairing as a more important mechanism. Particular musical patterns get played and replayed in specific contexts, gradually leading to a situation where exposure to that pattern automatically triggers associations with that dramatic content. Scholars working on eighteenth-century European art music have termed these standardized musical patterns that carry generally agreed upon associative referents *topics* (see Mirka, 2014). According to this account, musical features that align with patterns of conventional cultural and media usage should increase narrative engagement. For the sake of expediency, we call this dimension of musical features *topicality*. Excerpts high in topicality use patterns with conventional functions in mass media and culture. Invoking patterns with conventional functions makes it possible for brand-new music that a listener has never previously encountered to nevertheless evoke real world associations. Thus, topicality does not require widespread familiarity to an individual piece; rather,

a high topicality excerpt can be entirely unfamiliar but use standard patterns to which listeners have developed consistent extra-musical associations. For example, quiet descending arpeggios especially on the harp are often used to signal transportation into a sleep or dream state (Kim, 2018); particular drum beats are often used in the context of a military march (Norris, 2013). Note that as used here, topicality refers only to contemporary usage, not necessarily to compositional intent. For example, participants in Margulis (2017) generally agreed that an excerpt from the section marked Mephistopheles in Liszt's *Faust Symphony* triggered an imagined cat and mouse chase. The high topicality of this excerpt stemmed from its use of patterns that in contemporary media have been harnessed for Tom and Jerry cartoons, not from any intention on the part of Liszt to imply a plot line involving small cartoon animals. Proponents of associational pairing see music's social and cultural uses and contexts as a more important driver of narrative response than the intramusical characteristics of the sounds themselves, so topicality—a feature that requires exposure and enculturation—should drive narrative engagement.

In this article, two experiments investigate the relative contributions of contrast and topicality to narrative engagement with instrumental music. In Experiment 1, participants at two locations in the Midwest US (Arkansas and Michigan, where residents have ample exposure to Western media) and a rural village in China (Dimen, where residents have little exposure to Western media, and where residents speak Dong, a tone language) listened to short instrumental Western and Chinese excerpts that were High or Low in Topicality and Contrast, and which had previously been equated for enjoyment and familiarity across the four excerpt categories (High Topicality/High Contrast, Low Topicality/High Contrast, High Topicality/Low Contrast, Low Topicality/Low Contrast) at all research sites. They rated each excerpt using the Narrative Engagement (NE) Scale (Margulis et al., 2019), answered yes or no to a question about whether they imagined a story in response to the music (the Story Response Question or SRQ), and if yes, freely described any narrative they imagined. This article examines the NE and SRQ responses; the free response narratives are analyzed in a separate manuscript. NE and SRQ responses capture the extent to which any individual excerpt is heard narratively—whether (in the case of SRQ responses) and how absorbingly (in the case of NE scores) a listener imagined an ongoing narrative while listening. Narrative engagement is one possible mode of listening (cf. Tuuri & Eerola, 2012), and can be generated more

by individual excerpts than by others (Margulis et al., 2019). Narrative engagement is distinct from story content—two people could be imagining equally vivid stories of the same excerpt, but their stories could be entirely different. NE and SRQ measure how narratively individual excerpts are experienced, independent of the specific content of the stories.

If topicality generates narrative engagement, excerpts high on this attribute should elicit more within-culture narrative engagement. That is, reflecting distinct patterns of exposure between the research sites, High Topicality Western excerpts should elicit more narrative engagement at the US but not the Chinese research sites, and High Topicality Chinese excerpts should elicit more narrative engagement at the Chinese but not the US research sites. If, on the other hand, contrast is the primary generator of narrative engagement, excerpts high on this attribute should generate more narrative engagement regardless of cultural background. Since contrast is a musical feature that is relatively less dependent on enculturated exposure, it should operate similarly across cultures. Finally, if narrative engagement with music is not affected by topicality or by amount of contrast, there should be no significant difference in Narrative Engagement (NE) scores for excerpts featuring different levels of these parameters.

Experiment 2 extends Experiment 1 by more directly manipulating topicality and contrast. Participants in Experiment 2 completed the same narrative engagement task for original excerpts (that were either high in topicality or high in contrast, while low in the other feature) and for modified excerpts that reduced the attribute that had been high in the original version, so that the modified version was low in both topicality and contrast. For example, in a High Topicality excerpt that featured rapid descending arpeggios and was understood by many participants in previous studies to suggest a change in conscious state (e.g., falling asleep, dreaming), the Modified version retained all the details but blocked the chords (e.g., played them simultaneously instead of arpeggiating), removing the salient acoustic cue (descending arpeggios) thought to underlie the association. In a High Contrast excerpt that alternated between two sections with different instrumentations and dynamic levels the intervening contrasting section was removed, so that in the Modified version the instrumentation and dynamic levels remained stable throughout. Here, we were interested in whether directly attenuating these features could reduce the narrative engagement experienced by listeners. If reducing the topicality in High Topicality excerpts results in lower Narrative Engagement scores,

it suggests that topicality drives narrative response. If, on the other hand, reducing the contrast in High Contrast excerpts results in lower narrative engagement scores, it suggests that contrast drives narrative response. If, however, attenuating these features does not lower or increases Narrative Engagement scores, it weakens the case for their role in encouraging narrative response.

## Experiment 1

### METHOD

#### *Participants*

*Arkansas Group.* Three-hundred eighteen individuals (198 female), ages 17–30 years ( $M = 19.0$ ,  $SD = 1.6$ ) enrolled in a general psychology class participated in the experiment at the Music Cognition Lab at the University of Arkansas in Fayetteville, AR in exchange for partial course credit. Approximately 58% of participants reported no formal music training, defined as explicit instruction in music of any sort; the remaining 42% had between 1 and 14 years of music training ( $M = 5.1$ ,  $SD = 3.0$ ). Over 98% of participants reported watching English language media (currently:  $M = 11.8$  hours per week,  $SD = 12.9$ ; as a child:  $M = 15.9$  hours per week,  $SD = 14.0$ ). In contrast, only 15% of participants reported *ever* having been exposed to Chinese language media. For those 15%, current exposure to Chinese language media was estimated to be  $M = 1.4$  hours per week,  $SD = 2.5$ , while exposure as a child was estimated to be  $M = 3.4$  hours per week,  $SD = 4.4$ .

*Dimen Group.* One-hundred forty-seven individuals (126 female), ages 19–80 years ( $M = 46.6$ ,  $SD = 14.5$ ) from Dimen, China participated in the experiment at the Dimen Dong Community Cultural Research Center in Dimen, Guizhou Province, China. Approximately 70% of participants reported no formal music training; the remaining 30% of participants had between 1 and 60 years of music training ( $M = 14.5$ ,  $SD = 14.4$ ). Over 97% of participants reported watching Chinese language media (currently:  $M = 11.6$  hours per week,  $SD = 8.9$ ; as a child:  $M = 5.4$  hours per week,  $SD = 7.9$ ). In contrast, only 37% of participants reported *ever* having been exposed to English language media. For those 37%, current exposure to English language media was estimated to be  $M = 2.1$  hours per week,  $SD = 3.5$ , while exposure as a child was estimated to be  $M = 0.8$  hours per week,  $SD = 3.2$ .

*Michigan Group.* The within-culture comparison group to the Arkansas sample consisted of 157 individuals (120 female), ages 18–31 years ( $M = 19.2$ ,  $SD = 1.8$ ) enrolled in a general psychology class participated

in the experiment at the Timing, Attention, and Perception Lab at Michigan State University in Lansing, MI in exchange for partial course credit. Approximately 32% of participants reported no formal music training; the remaining 68% of participants reported between 1 and 18 years of music training ( $M = 5.9$ ,  $SD = 3.7$ ).

*Materials.* Stimuli were 60-s excerpts ( $n = 32$ ) drawn from commercial recordings of instrumental music with no lyrics or vocal part. Half of the excerpts ( $n = 16$ ) were drawn from recordings of Western art music and half ( $n = 16$ ) from recordings of Chinese art music. Preliminary pilot work determined that although participants in the Dimen group were broadly familiar with the style of Chinese music presented in the experiment, and participants in the Arkansas and Michigan groups were broadly familiar with the style of Western music presented in the experiment, these styles of music were not the ones to which participants tended to listen most frequently, and these specific excerpts were unlikely to be ones that participants had heard prior to the experimental session (i.e., they were relatively novel). Moreover, previous work using the same methods at the same three research sites (reported in Margulis et al., 2019) on a larger pool of 128 excerpts allowed for selecting excerpts matched for within-culture familiarity and enjoyment ratings, ensuring that these factors were equated across conditions.

Stimuli could be High or Low in both Topicality, a feature dependent on enculturation, and Contrast, a psychoacoustic feature less susceptible to enculturation, creating four stimulus categories in total (High Topicality, High Contrast; High Topicality, Low Contrast; Low Topicality, High Contrast; and Low Topicality, Low Contrast). Topicality assesses the degree to which individual excerpts make use of sonic patterns that feature within-culture conventionalized associations—for example horn calls associated with the hunt, or chromatic mediant progressions evocative of wonder (Cohn, 2012; Heine, 2018). Contrast assesses the degree to which individual excerpts vary along salient acoustic parameters.

Contrast and Topicality categories were assigned for each excerpt at an earlier stage of research. At this earlier stage, between 3 and 6 independent music theorists who were experts in either Western or Chinese art music rated the Topicality and musical Contrast of each of a large pool of Western excerpts or Chinese excerpts, respectively. Ratings were made on a 7-point scale (1 = *low*, 7 = *high*). Topicality was defined for raters as “the degree to which individual excerpts make

use of sonic patterns that feature within-culture conventionalized associations” and Contrast as “the degree to which individual excerpts vary along salient acoustic parameters.” Interclass correlation coefficients were calculated separately on Topicality and Contrast for Western and Chinese excerpts to assess the interrater reliability of the expert ratings. ICC values for Topicality and Contrast ratings for the Western excerpts were .64 and .82, respectively, indicating moderate to good reliability. ICC values for Topicality and Contrast ratings for the Chinese excerpts were .72 and .82, indicating good reliability. For this experiment, four excerpts in each of the four categories (HTHC, HTLC, LTHC, LTLC) were selected for the Western and Chinese excerpts, yielding 16 total Western excerpts and 16 total Chinese excerpts. Table 1 lists the source recordings for all stimuli used in the experiment.

*Procedure.* Once seated for the experiment, participants were instructed, “You’ll be asked to report aspects of your experience listening to musical excerpts, including whether or not you imagined a story while listening. Please do NOT specifically ATTEMPT to imagine a story. Simply listen to the music as you ordinarily would. If you imagine a story, that’s fine, and if you don’t imagine a story, that’s fine too.” Following these instructions, each participant heard one of four subsets of eight musical excerpts from the full set of 32—four Western and four Chinese with one excerpt for each category (HTHC, HTLC, LTHC, LTLC) for both Western and Chinese excerpts. Subset was rotated across participants so that NES and SRQ scores were obtained for all 32 excerpts with approximately an equal number of participants listening to each subset. Prior to the presentation of each excerpt, participants were told they should try to listen attentively as if they were intending to enjoy the piece. After listening to each excerpt, participants indicated whether they imagined a story or elements of a story while listening to the music (yes/no): the Story Response Question (SRQ). Next, they completed a four-item version of the Narrative Engagement (NE) scale. The NE scale consists of a four-statement inventory (“It was easy to imagine a story when listening to the music,” “I imagined a vivid story,” “I imagined a story with a clear setting, characters, and events,” and “I imagined a story while the music was playing, not afterwards”) to which participants respond on a 6-point scale (1 = *strongly disagree*, 6 = *strongly agree*) with a composite NE score obtained by averaging the responses to the four items. The procedures for developing the scale are reported in Margulis et al. (2019). The scale has been shown to exhibit

TABLE 1. Stimuli in Experiment 1

Title	Artist/Performer	Excerpt Type	Category (L=Low, H=High, C=Contrast, T=Topicality)
Phaéton, Op. 39	Camille Saint-Saëns, Lorin Maazel + Pittsburgh Symphony Orchestra	Western	LCLT
Pour le Piano L. 95 No. 3 Toccata	Claude Debussy, Claudio Abbado + Berlin Philharmonic	Western	LCLT
Das Lied Von Der Erde II. Der Einsame Im Herbst	Gustav Mahler, Michael Tilson Thomas + San Francisco Symphony	Western	LCLT
3 Pieces Op. 59 No. 3 Sonatina Pastorale	Sergei Prokofiev	Western	LCLT
March for Military Music in F Major “Yorck March” WoO 18	Ludwig van Beethoven, Hans Preim-Bergrath + Berlin Philharmonic Wind Ensemble	Western	LCHT
Grand Canyon Suite I. Sunrise	Ferde Grofé, Antal Doráti + Detroit Symphony Orchestra	Western	LCHT
6 Impromptus Op 5, No. 5 in B Minor	Jean Sibelius, Leif Ove Andsnes	Western	LCHT
In the Mystic Land of Egypt	Albert William Ketèlbey, Robert Sharples + New Symphony Orchestra	Western	LCHT
Four Etudes for Orchestra, No. 4: Madrid. Allegro con moto	Igor Stravinsky, Pierre Boulez + Chicago Symphony Orchestra	Western	HCLT
String Quartet No.5 - Part 3	Phillip Glass, The Smith Quartet	Western	HCLT
Suite from Dracula: Titles	Phillip Glass, Carducci Quartet	Western	HCLT
Piano Concerto No. 4 in G Major Op. 58, III. Rondo (Vivace)	Ludwig van Beethoven, Stephen Kovacevich + Colin Davis + London Symphony Orchestra	Western	HCLT
Symphony No. 2 “Resurrection”, I. Allegro maestoso. Mit durchaus ernstem und feierlichem Ausdruck	Gustav Mahler, David Zinman + Tonhalle Zürich	Western	HCHT
La Gazza Ladra: Overture	Gioachino Rossini, Gustavo Dudamel + Los Angeles Philharmonic	Western	HCHT
Billy the Kid Suite, II. Street in a Frontier Town	Aaron Copland, Leonard Bernstein + New York Philharmonic	Western	HCHT
Homenaje a Garcia Lorca: 1. Baiile	Silvestre Revueltas, Arthur Weisberg + Ensemble 21	Western	HCHT
Great Wave Washes the Sand	杨瑾 Yang Jin	Chinese	LCLT
Strains of Spring Morning	管平湖 Guan Pinghu	Chinese	LCLT
Level Sands, Lowering Geese	张维良 Zhang Weiliang	Chinese	LCLT
Timid Sound/Song	杨瑾 Yang Jin	Chinese	LCLT
Walking on an Ancient Road	赵良山 Zhao Liangshan	Chinese	LCHT
Bamboo Branch Lyric	群星 Various Artists	Chinese	LCHT
Farewell Pavilion, Slow Melancholy	赵良山 Zhao Liangshan	Chinese	LCHT
Groans of the Sick	周维 Zhou Wei	Chinese	LCHT
Empty Mountain, Birdsong	周维 Zhou Wei	Chinese	HCLT
Ballad Tune, Three-Six	张雪 Zhang Xue	Chinese	HCLT
The Great River Flows East	谭宝硕 Tan Baoshuo	Chinese	HCLT
Zhao Jun's Complaint	张雪 Zhang Xue	Chinese	HCLT
Ambush from Ten Sides	杨瑾 Yang Jin	Chinese	HCHT
Racing Horses	周维 Zhou Wei	Chinese	HCHT
Liuqin Opera Piece	群星 Various Artists	Chinese	HCHT
Spring Sun, White Snow	杨瑾 Yang Jin	Chinese	HCHT

Notes: Stimuli in Experiment 1 were divided between two excerpt types (Chinese or Western), two degrees of Topicality (High or Low), and two degrees of Contrast (High or Low), with four excerpts in each combination.

good internal consistency and validity. In addition to the four NE items, participants answered two questions pertaining to familiarity (whether they had heard the

specific excerpt before, and whether it sounded familiar to them) and two questions pertaining to pleasure (whether they enjoyed listening to the excerpt, and

whether they liked the excerpt using the same 6-point rating scale). After completing these items, participants responded to one of two free-response questions, depending on their response to the SRQ. If they answered yes to the SRQ, they described the story they imagined in as much detail as they were able. If they answered no to the SRQ, they were asked to speculate about why they did not imagine a story. Requesting free responses in both cases ensured that participants were not incentivized to select one option or the other simply on the basis of extent of subsequent task demands. Free response data are reported in a separate manuscript. Explicitly asking people whether they imagined stories may have primed participants to imagine more than they otherwise would; however, since this question was posed after every excerpt, it should affect every condition equally and not systematically bias the experiment's results.

Participants listened to the excerpts over high quality headphones with the presentation order of the eight excerpts randomized. Different listeners heard different sets of eight excerpts, so that narrative responses were obtained for all 32 excerpts across all participants. The entire experiment took approximately 50 minutes.

**Data Analysis.** The two primary dependent measures were the NE score (the average of the four NE items on the Narrative Engagement scale) and the SRQ score (the proportion of imagined story responses). Supplementary analyses also considered a familiarity score for each excerpt (the average of the two memory items on the Narrative Engagement scale) and an enjoyment score (the average of the enjoy and like items on the Narrative Engagement scale). All data analyses were conducted using SPSS version 26 with an alpha level of .05 as the criterion for significance.

## RESULTS

First, we examined the contributions of High vs. Low Topicality and High vs. Low Contrast to Narrative Engagement for the Midwest US listeners. If narrative responses to music arise from a fundamental tendency to structure abstract sounds in terms of stories, then excerpts with High Contrast should be the likeliest to be heard narratively regardless of culture; if, however, they arise from contextual exposure, then excerpts with High Topicality within a specific culture should be the likeliest to be heard narratively. Figure 1A-D shows Narrative Engagement (NE) results for both Michigan and Arkansas listeners for all conditions. Consistent with the enculturation hypothesis, a 2 (Culture of Excerpt: Western vs. Chinese) x 2 (Contrast: High vs.

Low) x 2 (Topicality: High vs. Low) ANOVA on NE scores for the Western listeners (Arkansas and Michigan combined) revealed a main effect of Culture of Excerpt,  $F(1, 24) = 9.85, p = .004, \eta^2 = 0.29$ , such that Western listeners were more narratively engaged by Western music ( $M = 3.81, 95\% \text{ CI} = 3.68\text{--}3.92$ ) than by Chinese music ( $M = 3.54, 95\% \text{ CI} = 3.42\text{--}3.66$ ), a main effect of Topicality,  $F(1, 24) = 7.08, p = .014, \eta^2 = 0.23$ , such that Western listeners were more narratively engaged by High Topicality excerpts ( $M = 3.79, 95\% \text{ CI} = 3.68\text{--}3.91$ ) than by Low Topicality excerpts ( $M = 3.56, 95\% \text{ CI} = 3.44\text{--}3.66$ ) and a significant interaction between type of excerpt and Topicality,  $F(1, 24) = 4.28, p = .049, \eta^2 = 0.15$ . The interaction revealed that High Topicality increased NE for Western excerpts (High Topicality,  $M = 4.01, 95\% \text{ CI} = 3.83\text{--}4.18$ ; Low Topicality,  $M = 3.61, 95\% \text{ CI} = 3.43\text{--}3.78, t(14) = 2.88, p = .01, \text{Cohen's } d = 1.44$ , but not for Chinese excerpts,  $t(14) = 0.46, p = .65, \text{Cohen's } d = 0.23$ , suggesting that Western listeners had received sufficient cultural exposure for the standard pattern associations to develop for Western music, but not for Chinese music.

Analysis of the SRQ scores yielded the same pattern of results as the analysis of the NE scores (see Figure 2). Consistent with the enculturation hypothesis, a 2 (Culture of Excerpt) x 2 (Contrast High vs. Low) x 2 (Topicality High vs. Low) ANOVA on Story Response Question (SRQ) scores for the Western listeners (Arkansas and Michigan combined) revealed a main effect of Culture of Excerpt,  $F(1, 24) = 13.43, p = .001, \eta^2 = 0.36$ , such that Western listeners heard more stories in response to Western music ( $M = 0.72, 95\% \text{ CI} = 0.69\text{--}0.75$ ) than to Chinese music ( $M = 0.64, 95\% \text{ CI} = 0.61\text{--}0.67$ ), a main effect of Topicality,  $F(1, 24) = 6.35, p = .02, \eta^2 = 0.21$ , such that Western listeners heard more stories in response to High Topicality excerpts ( $M = 0.71, 95\% \text{ CI} = 0.67\text{--}0.74$ ) than in response to Low Topicality excerpts ( $M = 0.65, 95\% \text{ CI} = 0.62\text{--}0.68$ ), and a significant interaction between Culture of Excerpt and Topicality,  $F(1, 24) = 5.58, p = .027, \eta^2 = 0.19$ , such that Topicality influenced SRQ scores for Western excerpts (High Topicality,  $M = 0.77, 95\% \text{ CI} = 0.73\text{--}0.82$ ; Low Topicality,  $M = 0.66, 95\% \text{ CI} = 0.62\text{--}0.71, t(14) = 3.49, p = .004, \text{Cohen's } d = 1.75$ , but not Chinese excerpts,  $t(14) = 0.11, p = .96, \text{Cohen's } d = 0.05$ .

Next, we considered contributions of familiarity and enjoyment to narrative engagement for the Midwest US listeners. Both familiarity and enjoyment were positively correlated with NE scores (familiarity,  $r = .66, p < .02$ ; enjoyment,  $r = .64, p < .01$ ). Nonetheless, the partial correlation between Topicality rating for each excerpt and NE score, controlling for familiarity and

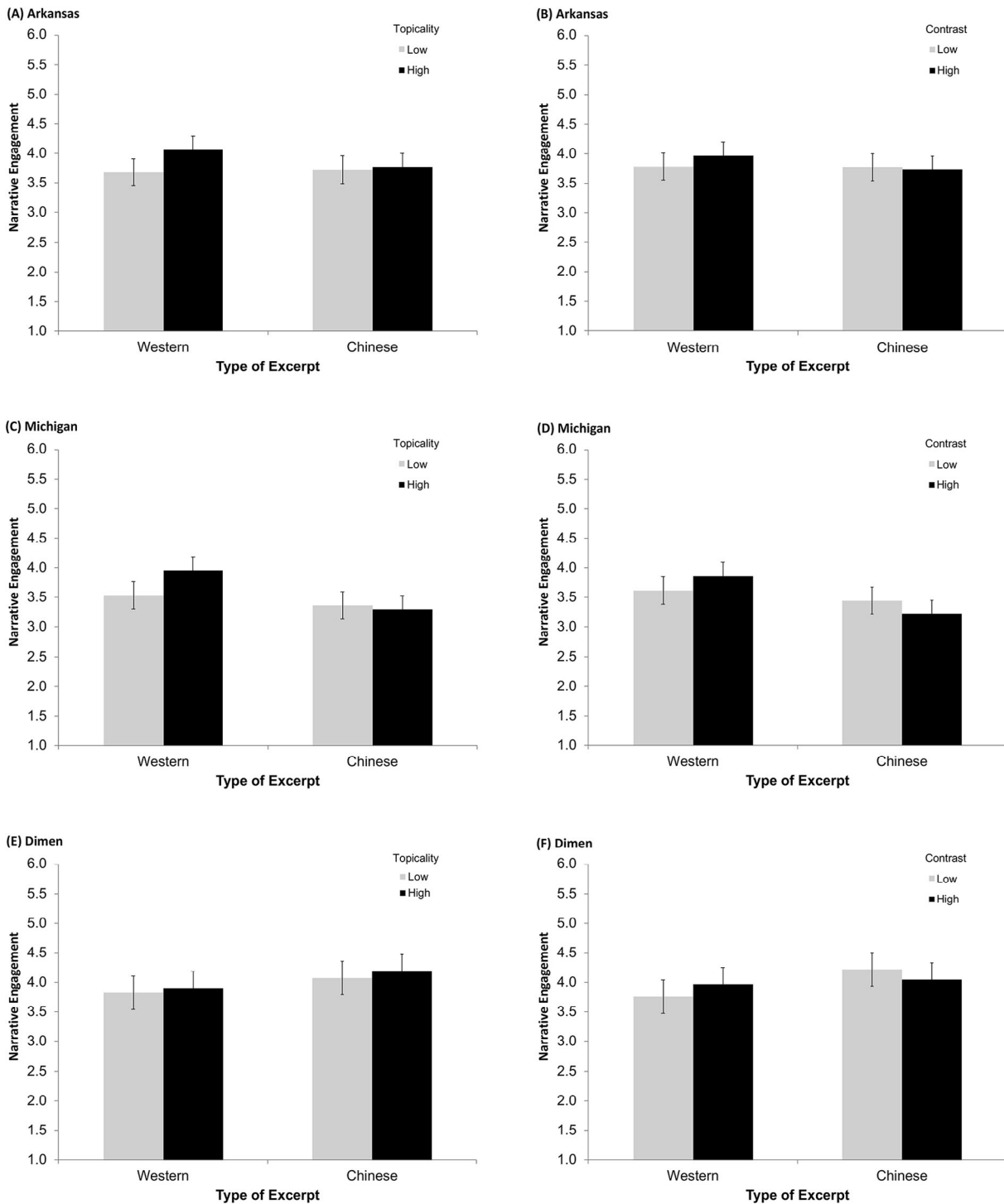
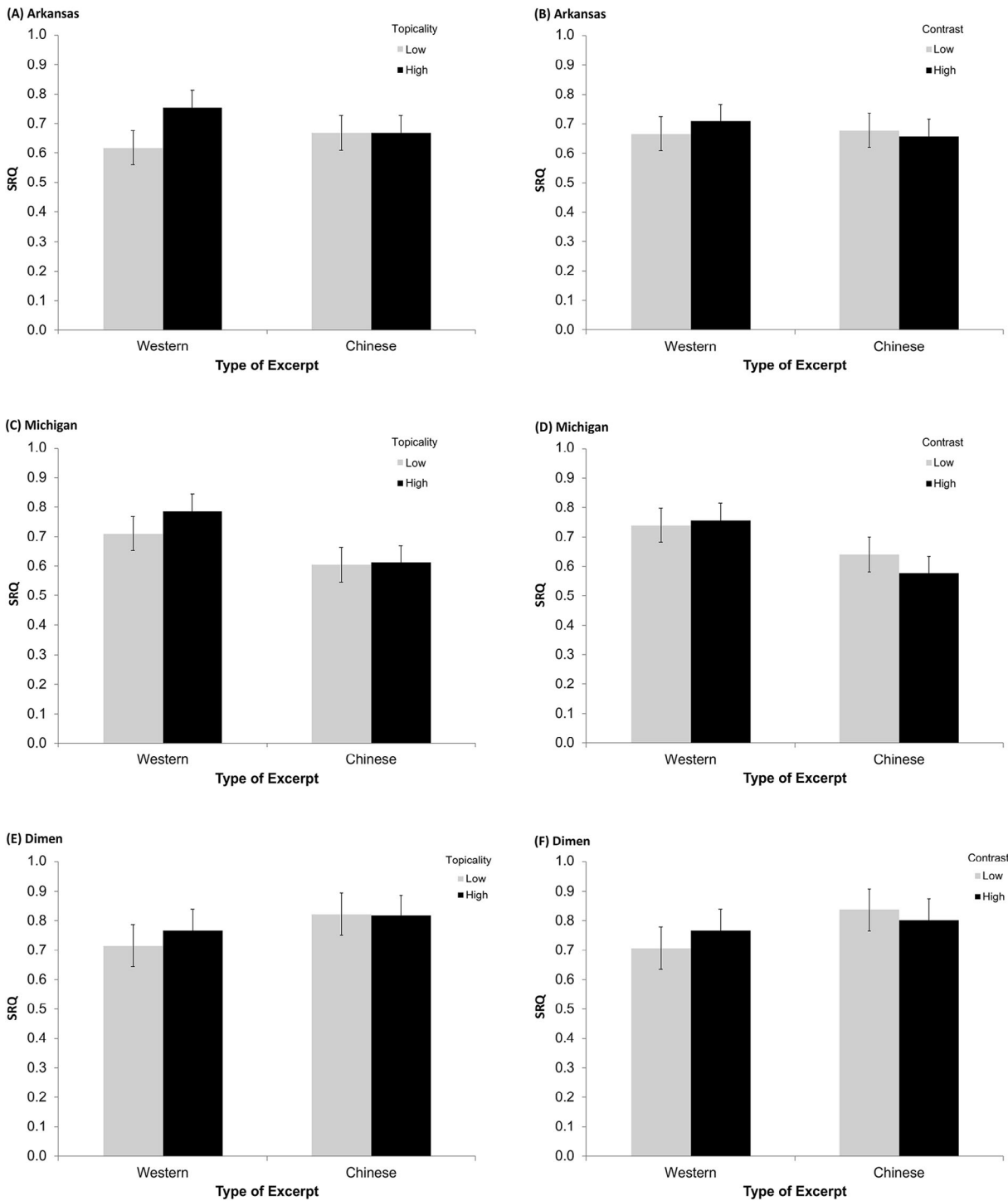


FIGURE 1. Narrative engagement scores for Arkansas (Panels A-B), Michigan (Panel C-D) and Dimen (Panels E-F) participants for Western excerpts and Chinese excerpts. The results for the Topicality manipulation are shown in the left column of panels (A, C, E), while the results for the Contrast manipulation are shown in the right column of panels (B, D, F). Error bars represent 95% confidence intervals. Both Arkansas and Michigan participants show greater narrative engagement for high topicality Western excerpts than for low topicality Western excerpts, but there was no effect of topicality for Chinese excerpts, and there was no effect of contrast for either the Western or Chinese excerpts. Dimen participants show greater NE for Chinese excerpts compared to Western excerpts, but show no effects of Topicality or Contrast.



**FIGURE 2.** SRQ scores for Arkansas (Panels A-B), Michigan (Panel C-D) and Dimen (Panels E-F) participants for Western excerpts and Chinese excerpts. The results for the Topicality manipulation are shown in the left column of panels (A, C, E) while the results for the Contrast manipulation are shown in the right column of panels (B, D, F). Error bars represent 95% confidence intervals. Both Arkansas and Michigan participants heard more stories for high topicality Western excerpts than for low topicality Western excerpts, but there was no effect of Topicality for Chinese excerpts, and there was no effect of Contrast for either the Western or Chinese excerpts. Dimen participants show greater NE for Chinese excerpts compared to Western excerpts, but show no effects of Topicality or Contrast.



enjoyment, was still significant ( $r = .41$ ,  $p = .026$ ). A stepwise regression revealed that Topicality, familiarity and enjoyment independently contribute to NE scores and together account for 58% of the variance. Inclusion of Contrast ratings as a predictor does not contribute additional variance.

Next, we examined the contributions of High vs. Low Topicality and High vs. Low Contrast to Narrative Engagement for the Dimen listeners. Figure 1E-F shows that Dimen listeners, opposite to the Midwest US participants, perceive Chinese excerpts slightly more narratively ( $M = 4.13$ , 95% CI = 4.03–4.27) than Western excerpts ( $M = 3.86$ , 95% CI = 3.70–4.01),  $F(1, 24) = 3.73$ ,  $p = .05$ ,  $\eta^2 = 0.13$ . Moreover, there are no main effects of Topicality or Contrast, or reliable interactions (all  $p$ 's > .17). The same 2 x 2 x 2 ANOVA on SRQ scores for Dimen listeners revealed the same main effect of culture of excerpt,  $F(1, 24) = 4.99$ ,  $p = .035$ ,  $\eta^2 = 0.17$ , such that the proportion of heard stories was greater for Chinese excerpts ( $M = 0.82$ , 95% CI = 0.79–0.85) than for Western excerpts ( $M = 0.74$ , 95% CI = 0.70–0.78); see Figure 2E-F. Identical to the analysis of NE scores, there were no other main effects or interactions for SRQ scores (all  $p$ 's > .13).

Finally, we considered contributions of familiarity and enjoyment to narrative engagement for the Dimen listeners. Both familiarity and enjoyment were positively correlated with NE scores (familiarity,  $r = .82$ ,  $p < .001$ ; enjoyment,  $r = .37$ ,  $p < .05$ ). Moreover, consistent with the ANOVAs, neither Topicality ratings, nor Contrast ratings, for each excerpt, were correlated with NE (topicality,  $r = .14$ ,  $p = .45$ ; contrast,  $r = .17$ ,  $p = .36$ ). A stepwise regression with Topicality rating, Contrast rating, familiarity, and enjoyment as predictors showed that familiarity alone accounts for 67% of the variance in the NE scores for Dimen listeners, and that none of the other predictors contributes additional unique variance.

## Discussion

In sum, results show that participants at US sites exhibit greater narrative engagement (as measured by NE scores, as well as by SRQ) to Western excerpts than to Chinese excerpts and greater narrative engagement for Western excerpts with High Topicality than with Low Topicality. The comparison of High vs. Low Contrast excerpts (at least as measured within our study) had no significant effect on NE or SRQ scores for either Western or Chinese excerpts. Familiarity and enjoyment also contribute to Narrative Engagement, but these contributions are distinct from Topicality. Dimen

listeners in contrast showed an effect of excerpt culture: Chinese excerpts were more narratively engaging than Western excerpts, but excerpt familiarity and enjoyment were the only other factors that affected narrative engagement.

The null effect of Topicality for the Western excerpts in Dimen compared to the significant effect of Topicality for the Western excerpts in Arkansas and Michigan is likely related to the comparatively lower exposure to Western media in Dimen, where the majority of participants reported *never* having been exposed to it. Dimen participants did, however, have some exposure to Chinese media, yet a significant effect of Topicality did not arise for those excerpts either. Since the media that Dimen participants access is typically in Mandarin or Chinese, and the majority of participants speak only Dong, it is possible that the language gap interfered with the generation of associations. It is also possible that broader cultural experiences beyond exposure to mass media are critical to the generation of associational pairings. As members of one of 55 officially recognized minority groups in China, participants in Dimen share not only a different language from the Han minority, but also a distinct set of day-to-day experiences and cultural exposures.

Experiment 2 investigated whether we could experimentally manipulate narrative responses by reducing Topicality and Contrast to Low in individual excerpts that started out High along one of these parameters and having people perform the same tasks from Experiment 1 on both the original and manipulated excerpts, providing causal evidence for the role of Topicality.

## Experiment 2

### METHOD

**Participants.** Eighty-seven individuals (50 female), ages 18–22 years ( $M = 19.5$ ,  $SD = 0.8$ ), enrolled in a general psychology course participated in the experiment at the Music Cognition Lab at the University of Arkansas in Fayetteville, AR in return for partial course credit in an undergraduate psychology course. Approximately 59% of participants reported no formal music training; the remaining 41% reported between 1 and 15 years of music training ( $M = 6.1$ ,  $SD = 3.8$ ). Over 96% of participants reported watching English language media (currently:  $M = 14.4$  hours per week,  $SD = 14.5$ ; as a child:  $M = 17.3$  hours per week,  $SD = 15.8$ ).

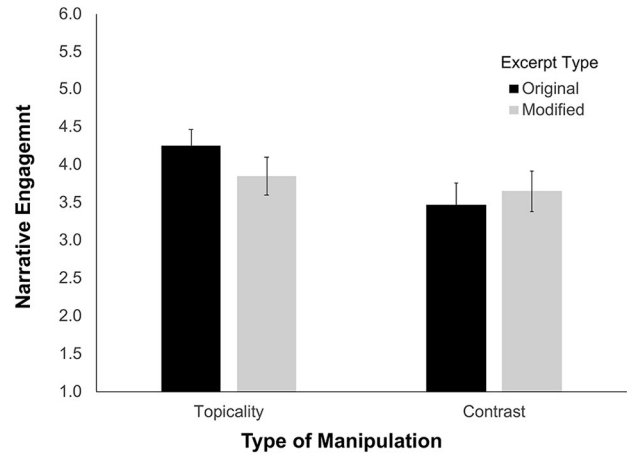
**Materials.** Stimuli were 12 one-minute excerpts, six drawn from stimuli used in previous studies (the Original versions), and six consisting of modified versions of the stimuli (the Modified versions). Of the six Original

versions, three were High Topicality, Low Contrast stimuli, and three were High Contrast, Low Topicality stimuli (see description above). The Modified versions removed salient acoustic cues for Topicality (in the case of the three High Topicality excerpts) or for Contrast (in the case of the three High Contrast excerpts). For example, in a High Topicality excerpt that featured rapid descending arpeggios and was understood by many participants in previous studies to suggest a change in conscious state (e.g., falling asleep, dreaming), the Modified version retained all the details but blocked the chords (e.g., played them simultaneously instead of arpeggiating), removing the salient acoustic cue (descending arpeggios) thought to underlie the association, changing the Topicality parameter from High to Low from the Original to the Modified Version. In an excerpt that alternated between two sections with different instrumentations and dynamic levels, the intervening contrasting section was removed so that the instrumentation and dynamic levels remained stable throughout, changing the Contrast parameter from High to Low from the Original to the Modified Version.

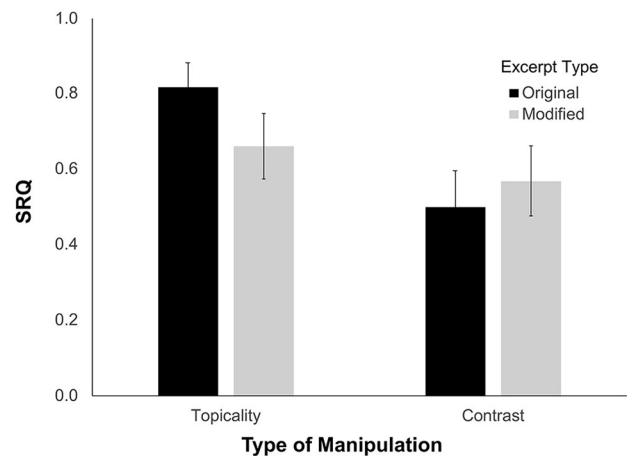
**Procedure.** The procedures were identical to Experiment 1, except that each Arkansas participant heard 6 of the 12 one-minute excerpts, each in either Original or Modified version. No participant heard the same excerpt in both versions (Original and Modified) and each participant heard half of the stimuli in Original and half in Modified form.

## Results and Discussion

Figure 3 shows Narrative Engagement (NE) scores for Original and Modified excerpts for excerpts that were originally High in Topicality and Low in Contrast (left pair of bars) and those that were originally High in Contrast and Low in Topicality (right pair of bars). A 2 (Type of Modification: Topicality vs. Contrast)  $\times$  2 (Type of Excerpt: Original vs. Modified) repeated-measures ANOVA on NE scores revealed a main effect of Type of Modification,  $F(1, 86) = 21.8, p < .001, \eta^2 = 0.20$ , and a significant interaction between Type of Modification and Type of Excerpt,  $F(1, 86) = 5.41, p = .02, \eta^2 = 0.06$ . Results for the Story Response Question (SRQ) scores paralleled those found for the NE scores (see Figure 4). A 2 (Type of Modification: Topicality vs. Contrast)  $\times$  2 (Type of Excerpt: Original vs. Modified) repeated-measures ANOVA on SRQ scores revealed a main effect of Type of Modification,  $F(1, 86) = 29.83, p < .001, \eta^2 = 0.26$ , and a significant interaction between Type of Modification and Type of Excerpt,  $F(1, 86) = 7.54, p = .007, \eta^2 = 0.08$ . Consistent



**FIGURE 3.** Narrative engagement (NE) scores with 95% confidence intervals for original and modified musical excerpts that were originally high in topicality and low in contrast (left pair of bars) compared with those that were originally high in contrast and low in topicality (right pair of bars). Excerpts modified to reduce topicality generated lower narrative engagement than the original excerpts. The manipulation of contrast, on the other hand, had no effect on either the number of heard stories or the level of narrative engagement



**FIGURE 4.** Proportion of imagined stories with 95% confidence intervals for original and modified musical excerpts that were originally high in topicality and low in contrast (left pair of bars) compared with those that were originally high in contrast and low in topicality (right pair of bars). Excerpts modified to reduce topicality generated fewer heard stories than the original excerpts. The manipulation of contrast, on the other hand, had no effect on either the number of heard stories.

with predictions, excerpts modified to reduce Topicality received lower Narrative Engagement scores (Modified,  $M = 3.86, 95\% \text{ CI} = 3.60\text{--}4.11$ ; Original,  $M = 4.26, 95\% \text{ CI} = 4.05\text{--}4.47$ ),  $t(86) = -2.65, p = .01$ , Cohen's  $d = -0.28$ , and resulted in fewer heard stories (Modified,

$M = 0.66$ , 95% CI = 0.57–0.75; Original,  $M = 0.82$ , 95% CI = 0.76–0.88),  $t(86) = -2.73$ ,  $p = .008$ , Cohen's  $d = -0.30$ , than the original excerpts. In contrast, excerpts modified to reduce Contrast did not differ in their NE scores,  $t(86) = 0.92$ ,  $p = .36$ , Cohen's  $d = 0.10$ , or in the proportion of heard stories,  $t(86) = 1.01$ ,  $p = .31$ , Cohen's  $d = 0.11$ , compared to the original excerpt. In sum, in line with Experiment 1, explicitly reducing Topicality reduced NE and SRQ scores, but reducing Contrast did not, further reinforcing the central role of association and enculturation in narrative experiences of music.

### General Discussion

Two experiments investigated the question of what drives narrative engagement with music. A feature heavily dependent on culture and exposure, Topicality, but not a less experience-dependent feature, Contrast, was found to affect Western listeners' narrative engagement with Western excerpts, but not Chinese excerpts, as would be expected if enculturated associations to culturally familiar musical styles drives narrative perceptions of music. Excerpts with higher Topicality elicited higher narrative engagement in listeners with sufficient enculturation to those styles of music, consistent with the notion that imagined stories arise from exposure to repeated linkages between sound patterns and associated imagery, environments, and plotlines in multimedia contexts. This effect of topicality held when controlling for excerpt enjoyment and familiarity, which made separate, but independent contributions to narrative engagement. The capacity of musical patterns to accrue narrative implications across the course of cultural embedding and exposures underscores the fundamental multimodality of musical experience (Eitan, 2017).

This topicality effect occurred in listeners at both US research sites, but not at the Dimen research site. The absence of a topicality effect in Dimen for Western excerpts parallels the absence of a topicality effect in Arkansas and Michigan for Chinese excerpts, and possibly stems from the same source: insufficient exposure to the musical patterns in relevant cultural and media contexts. The absence of a topicality effect in Dimen for Chinese excerpts is more puzzling, given that the Dimen participants have access to Chinese media, exposure to which might have been sufficient for associations to arise. However, at least two other factors might have interfered. Given that mainstream Chinese movies and television shows are typically in Mandarin or Cantonese, languages that the Dong participants do not speak (Dong belongs to the Tai family, which is substantively different from languages in the Sino-

Tibetan family, including Mandarin; Ramsey, 1989), it is possible that this language barrier interfered with the formation of associations that might have been typical for Mandarin speakers. Furthermore, it is likely that topicality effects depend on exposure not just to mass media but also to other kinds of everyday cultural contexts. As members of the Dong community, one of China's 55 officially recognized ethnic minorities, Dimen participants' daily life in their village entails experiences and exposures that diverge markedly from those of the majority Han Chinese. Indeed, the free response narratives provided by Dimen participants (analyzed in a separate manuscript) support this supposition. Although their free response stories exhibit substantial across-participant agreement in response to individual excerpts, these stories often depart substantially from the typical associations of the same excerpts within Han Chinese culture.

Regardless, future work should further explore the lack of a Topicality effect in Dimen participants. Toward this end, one potentially fruitful line of new investigation would be to consider in more detail the one factor that did predict Dimen participants' by-excerpt narrative engagement—participants' familiarity with the excerpt. Given that topicality represents the formation of consistent associations between musical patterns and extramusical events, it is possible that the link observed between familiarity and narrative engagement for the Dimen participants represents a precursor to Topicality playing a more prominent role. The formation of abstract associations between musical patterns and extramusical events may require familiarity with a sufficient number of exemplar excerpts where listeners experience that musical pattern in a particular extramusical context. Immersed in their own Dong musical traditions, Dimen listeners may simply not have reached that threshold for either Western or Chinese music.

Across all sites, Contrast, an acoustic feature that should impact narrativization in a way that is less dependent on enculturation, showed little evidence of driving narrative response. This tends to support the idea that a person's "listening biography" (Wong, Chan, Roy, & Margulis, 2011)—their prior, contextually situated experiences with music—plays a larger role in generating semantic type responses than intrinsic properties of the sound itself. The comparatively larger role of experience over acoustic features is supported by previous work's finding that the excerpts that Dimen participants narratively engage to most are unrelated to the excerpts that Arkansas and Michigan participants narratively engage to most (Margulis et al., 2019). If stimulus properties drive narrative response

independent of cultural exposure and experience, then the same set of excerpts should have elicited high narrative engagement at every research site; instead, Margulis et al. (2019) found that different sets of excerpts elicited high narrative engagement in Dimen versus the US sites. Together with this paper's finding that topicality—a feature that can only impact narrative engagement if participants have specific kinds of cultural experiences—sometimes impacted narrative response, but contrast—a feature theorized to impact narrative engagement independent of listener enculturation—never did, this suggests that experience and exposure may play an especially significant role. This possibility raises a potential problem for music psychology, which often relies on varying properties of the acoustic signal more than varying aspects of the previous experiences participants bring to the listening session. A recent increase in cross-cultural studies within music cognition (Stevens, 2012) about not only high-level but also low-level perceptual phenomena (Jacoby & McDermott, 2017) has the potential to address this shortcoming and help awaken the field to more of the particulars and variability that define musical experiences. Nevertheless, as the complexities of this study demonstrate, cross-cultural work brings multiple challenges as well as potential pitfalls (see Jacoby et al., 2020). Future work with the Dimen participants would benefit from the inclusion of Dong individuals on the research team, so that hypotheses and methods could be developed from the perspective of a cultural insider.

It is also important to note that other stimulus features not tested within this study may have also played a role. Even though overall ratings of musical contrast seemed not appear to drive narrative engagement for this set of excerpts, it's possible that contrast might have a larger effect for other excerpts. We examined overall ratings of musical contrast because of its centrality to music theoretic accounts of narrative, but it's possible that more specific contrasting elements related to different musical characteristics (such as tempo, loudness, key, and articulation) that are not as experience-dependent as topicality contribute to narrative perceptions of music.

In this paper, the role of topicality for within-culture excerpts was demonstrated both by studying responses to commercially available recordings, and by experimentally manipulating these recordings to change the feature of interest. Future work could use specially composed excerpts to further test the role of topicality and connect it to practical questions in music composition and generation. Excerpts of production music, for example, generate broadly consistent imagery in listeners (Huovinen & Kaila, 2015), pointing to the

essential but largely neglected by scholars role of semantics in musical experience (see also McAuley et al., 2021). An important undertaking for future work is to probe topicality in contemporary listening with the aim of excavating key patterns and their associations within specific populations. Scholars have done this quite thoroughly for topics in eighteenth-century music as understood by this music's contemporaries (cf. Mirka, 2014) but the understandings of present day listeners have been comparatively understudied, at least by music theory and music psychology.

This study also points to narrative engagement as a readily available mode of listening to music. Future studies can track how these narrative imaginings relate to other dynamic and valenced musical responses, such as tension perceptions. Given that narrative engagement tracks closely with enjoyment and interest (Margulis et al., 2019), it is possible that imagined stories operate as a latent variable, influencing responses in common experimental paradigms in ways that have not yet been acknowledged. Additionally, it is likely that the pool of media and cultural exposure, as well as the relevant cognitive processing (see Salthouse, 2019) varies between people in different age brackets, even at one research site—an issue that also could have affected results in the present study, given that participants at the US research sites were significantly younger than participants at the Dimen research site. A future line of research compares responses on this task between listeners aged 18–22 and listeners aged 65–69 at the same US research site—a cross-cultural manipulation that operates along a different axis than geography.

Culturally situated exposure drives associations that feel natural and inevitable, yet are in fact contingent on enculturated pairings. The sense that music's expressive associations exist within the music itself when they in fact depend on specific sets of experiences and exposure reveal music's power, potential, and danger as a communicative force. Stories that seem to arise without mediation from the sounds themselves in fact depend on complex patterns of cultural exposure. The fact that music's affective and associational force often feels so immanent makes it a potent force for both connection and division across people and cultures, suggesting that a broader understanding of the associational contingency of musical expressivity could aid intercultural understanding.

#### Author Note

This research was supported by the Division of Behavioral and Cognitive Sciences of the National Science Foundation, Award Numbers 1734025 (PI: EHM) and

1734063 (PI: JDM). Xin Kang, Jieqiong Che, Xiyu Wang, Xueying Xu, Zhentian Liu, Chunzi Li, Xiaotong Ge, and Shengnan Zhao helped with data collection and translation for Dimen participants. Rhimmon Simchy-Gross and Lauren Shepherd helped with data collection in Arkansas. Many thanks to Anusha Mamidipaka, Gabby Kindig, and Jewelina Fairchild for their assistance with data collection and to the members of the Timing, Attention and Perception Lab at Michigan State University for their many helpful comments. Special thanks to Mr. LEE Wai Kit and

the staff at the Dimen Dong Eco-Museum for making data collection possible, and to the people in Dimen who participated in this research.

*Correspondence concerning this article should be addressed to* Address correspondence to: Elizabeth Hellmuth Margulis, Department of Music, Princeton University, Princeton, NJ 08544. E-mail: margulis@princeton.edu; J. Devin McAuley, Department of Psychology, Michigan State University, East Lansing, MI 48840. E-mail: dmcauley@msu.edu

## References

- ALMÉN, B. (2008). *A theory of musical narrative*. Indiana University Press.
- COHN, R. (2012). Peter, the wolf, and the hexatonic uncanny. In F. Wörner, U. Scheidler, & P. Rupprecht (Eds.), *Tonality 1900-1950: Concept and Practice* (pp. 47–62). Franz Steiner Verlag.
- EITAN, Z. (2017). Musical connections: Cross-modal correspondences. In R. Ashley & R. Timmers (Eds.), *The Routledge companion to music cognition* (pp 213–224). Routledge.
- FELD, S. (1984). Sound structure as social structure. *Ethnomusicology*, 28, 383–409.
- HEIDER, F., & SIMMEL, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, 57, 243–259.
- HEINE, E. (2018). Chromatic mediant and narrative context in film. *Music Analysis*, 37, 103–132.
- HERBERT, R. (2011). *Everyday music listening: Absorption, dissociation, and trancing*. Ashgate.
- HUOVINEN, E., & KAILA, A. K. (2015). The semantics of musical topoi. *Music Perception*, 33, 217–243.
- JACOBY, N., MARGULIS, E. H., CLAYTON, M., HANNON, E., HONING, H., IVERSEN, J., ET AL. (2020). Cross-cultural work in music cognition: Challenges, insights and recommendations. *Music Perception*, 37, 185–195.
- JACOBY, N., & McDERMOTT, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Current Biology*, 27, 359–370.
- KASSABIAN, A. (2001). *Hearing film: Tracking identifications in contemporary Hollywood film music*. Routledge.
- KIM, J. (2018). Change and transformation: The harp as a symbol of liminality in Tchaikovsky's *The Nutcracker* (1892) (Unpublished master's thesis). San Jose State University.
- KÜSSNER, M. B., & EEROLA, T. (2019). The content and functions of vivid and soothing visual imagery during music listening: Findings from a survey study. *Psychomusicology: Music, Mind, and Brain*, 29(2–3), 90–99.
- MARGULIS, E. H. (2017). An exploratory study of narrative experiences of music. *Music Perception*, 35, 235–248.
- MARGULIS, E. H., WONG, P. C. M., SIMCHY-GROSS, R., & MCAULEY, J. D. (2019). What the music said: Narrative listening across cultures. *Palgrave Communications*, 5, 146.
- MCAULEY, J. D., WONG, P. C. M., MAMIDIPAKA, A., PHILLIPS, N., & MARGULIS, E. H. (2021). Do you hear what I hear? Perceived narrative constitutes a semantic dimension for music. *Cognition*, 212, 104712.
- MIRKA, D. (2014). *The Oxford handbook of topic theory*. Oxford University Press.
- NORRIS, J. (2013). *Marching to the drums: A history of military drums and drummers*. Spellmount.
- PRESICCE, G., & BAILES, F. (2019). Engagement and visual imagery in music listening: An exploratory study. *Psychomusicology: Music, Mind, and Brain*, 29(2-3), 136–155.
- RAMSEY, S. R. (1989.) *The languages of China*. Princeton University Press.
- SALTHOUSE, T. A. (2019). Trajectories of normal cognitive aging. *Psychology and Aging*, 34, 17–24.
- STEVENS, C. J. (2012). Music perception and cognition: A review of recent cross-cultural research. *Topics in Cognitive Science*, 4, 653–667.
- TAGG, P., & CLARIDA, B. (2003). *Ten little title tunes: Towards a musicology of the mass media*. The Mass Media Music Scholars' Press.
- TAN, S.-L., COHEN, A. J., LIPSCOMB, S. D., & KENDALL, R. A. (2013). *The psychology of music in multimedia*. Oxford University Press.
- TUURI, K., & EEROLA, T. (2012). Formulating a revised taxonomy for modes of listening. *Journal of New Music Research*, 41(2), 137–152.
- VROEGH, T. (2019). Zoning-in or tuning-in? Identifying distinct absorption states in response to music. *Psychomusicology: Music, Mind, and Brain*, 29(2–3), 156–170.
- WONG, P. C. M., CHAN, A. H. D., ROY, A., & MARGULIS, E. H. (2011). The bimusical brain is not two monomusical brains in one: Evidence from musical affective processing. *Journal of Cognitive Neuroscience*, 23, 4082–4093.