

# Improvisation experience predicts how musicians categorize musical structures

Psychology of Music

2020, Vol. 48(1) 18–34

© The Author(s) 2018

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0305735618779444

journals.sagepub.com/home/pom



Andrew Goldman<sup>1,2</sup>, Tyreek Jackson<sup>3,4</sup>  
and Paul Sajda<sup>1,5</sup>

## Abstract

Western music improvisers learn to realize chord symbols in multiple ways according to functional classifications, and practice making substitutions of these realizations accordingly. In contrast, Western classical musicians read music that specifies particular realizations so that they rarely make such functional substitutions. We advance a theory that experienced improvisers more readily perceive musical structures with similar functions as sounding similar by virtue of this categorization, and that this categorization partly enables the ability to improvise by allowing performers to make substitutions. We tested this with an oddball task while recording electroencephalography. In the task, a repeating standard chord progression was randomly interspersed with two kinds of deviants: one in which one of the chords was substituted with a chord from the same functional class (“exemplar deviant”), and one in which the substitution was outside the functional class (“function deviant”). For function compared to exemplar deviants, participants with more improvisation experience responded more quickly and accurately and had more discriminable N2c and P3b ERP components. Further, N2c and P3b signal discriminability predicted participants’ behavioral ability to discriminate the stimuli. Our research contributes to the cognitive science of creativity through identifying differences in knowledge organization as a trait that facilitates creative ability.

## Keywords

*Improvisation, concepts and categories, creativity, EEG, auditory perception/cognition*

Many human behaviors can be characterized as improvisatory: past knowledge allows us, in real time, to understand and respond to a wide range of possible situations that we may never have exactly encountered before. Improvisation may seem esoteric—the term most often being

<sup>1</sup>Laboratory for Intelligent Imaging and Neural Computing, Department of Biomedical Engineering, Columbia University, USA

<sup>2</sup>Presidential Scholar in Society and Neuroscience, Columbia University, USA

<sup>3</sup>Department of Biobehavioral Sciences, Teachers College, Columbia University, USA

<sup>4</sup>Department of Music and Music Education, Teachers College, Columbia University, USA

<sup>5</sup>Data Science Institute, Columbia University, USA

## Corresponding author:

Andrew Goldman, Presidential Scholar in Society and Neuroscience, Columbia University, 511 Fayerweather Hall, MC 2509, 1180 Amsterdam Avenue, New York, NY 10027, USA.

Email: [ajg2232@columbia.edu](mailto:ajg2232@columbia.edu)

applied to musicians—but the concept has been used to describe a wide variety of human activities (Lewis & Piekut, 2016). For the purposes of cognitive-scientific research, a better definition of the construct and according methods to measure and explain its mechanisms would contribute to a better understanding of the creative mind.

Western musical practices, more so than other improvisatory behaviors like conversing and navigating, allow improvisatory ability to be isolated. People who know a language are almost always able to generate novel utterances in that language. However, there are many musicians with considerable expertise and deep understanding of musical styles who cannot generate new “utterances,” while others with similar levels of training (albeit different kinds of training) *are* able to improvise. To understand this difference is to understand what ways of learning and ways of knowing facilitate the ability to use that knowledge creatively.

Past neuroscientific studies on musical improvisation (Bengtsson, Csíkszentmihályi, & Ullén, 2007; Berkowitz & Ansari, 2008, 2010; de Manzano & Ullén, 2012; Donnay, Rankin, Lopez-Gonzalez, Jiradejvong, & Limb, 2014; Limb & Braun, 2008; Liu et al., 2012) have been concerned with what brain states underlie the spontaneous generation of novel motor behavior. Other studies have taken a formalistic perspective (i.e., examine the structural characteristics of the sounds produced by performers) and make inferences about the generative algorithms they use (Johnson-Laird, 1991, 2002), or psychological and cognitive processes they use (Frieler, Pflöderer, Zaddach, & Abeßer, 2016; Goldman, 2013; Hargreaves, 2012; Norgaard, 2011, 2014; Norgaard, Spencer, & Montiel, 2013; Pflöderer & Frieler, 2010; Pressing, 1988, 1998). Essentially, these studies isolate and describe improvisatory states and processes and describe the different kinds of knowledge improvisers use.

Another approach is to examine *traits* of experienced versus inexperienced improvisers that arise through differences in training and experience. Differences in sensitivity to changes in various musical features (Tervaniemi, Janhunen, Kruck, Putkinen, & Huotilainen, 2015; Vuust et al., 2005; Vuust, Brattico, Seppänen, Näätänen, & Tervaniemi, 2012), spatial reasoning (Harris & de Jong, 2015), and functional connectivity in the brain (Pinho, de Manzano, Fransson, Eriksson, & Ullén, 2014) have been observed between these groups. The present study builds on these between-group comparison approaches by providing another hypothesis as to how improvisers' and non-improvisers' cognitive processes may differ, and, importantly, the theoretical framework to help explain why such a difference would help enable the ability to improvise.

### *Differences in the use of musical structures*

Western improvising musicians in tonal contexts often use notation that does not fully specify which notes to play (e.g., figured bass and lead sheets); rather, symbols specify classes of harmonies and melodies. The performer chooses between a set of possible realizations within the functional category specified by the symbol. By function, we mean that different exemplars can be *used* similarly according to both explicit music theoretical rules or through implicit knowledge of the regularities of a style. For an improviser, such an ability to switch fluently between exemplars in performance would enable, for instance, harmonizing a melody in a new way, or soloing over a familiar harmonic progression with a performance partner who is using different exemplars of that progression. By contrast, practitioners of Western classical art music—who rarely improvise harmonic and melodic aspects of music—typically use notation that fully specifies which notes to play. In a given piece, a classical musician could, in principle, substitute certain musical structures for other similarly functioning ones, but they rarely do this as it contradicts the practice and tradition of classical music, which treats music as fixed works (Goehr, 1992).

Hypothetically, for improvisers, two exemplars within the same functional category would be perceived as more similar than a chord outside the functional category. For less experienced improvisers, this difference should be less pronounced since these kinds of exemplar substitutions are rarely made in practice, thus not reinforcing musical function as a categorization criterion. This research question is formulated with tonal improvisation in mind, although similar ideas may apply to other traditions of improvisation (see Discussion).

### *The current study*

We tested these differences in categorization using an oddball paradigm while recording electroencephalography (EEG). Our musical stimuli allowed us to compare the relative salience of different kinds of harmonic changes we termed “exemplar deviants” (changes within a functional category) and “function deviants” (changes between functional categories) in a sample of musicians with varying levels of improvisatory experience. For example, if the standard chord progression was ii–V–I, an exemplar deviant would be ii–V<sup>6</sup>–I and a function deviant would be ii–IV<sup>6</sup>–I. In addition to harmonic differences, such changes also necessarily change the intervals between specific notes, which could also change the relative salience of each kind of deviant; we controlled for these changes (see Methods). Improvisation experience should lead to categorizing exemplar deviants as being more similar to standards than function deviants are; behaviorally, function deviants should thus be identified more quickly and accurately than exemplar deviants, dependent on the amount of improvisation experience.

EEG data allowed us to assess the neurophysiological activity that discriminated each kind of deviant stimulus from the standard. Specifically, in an oddball task, if detection of a deviant requires classification (e.g., detecting deviant male names among standard female names) rather than detecting physical differences in stimuli only, an N2c component is evoked (Ritter, Simson, & Vaughan, 1983). This requires a controlled process rather than a purely stimulus driven one. Although other N2 subcomponents—namely the mismatch negativity (MMN) and the N2b—have similar latencies and topographies (for auditory stimuli), we interpret our results in terms of the N2c due to the nature of the task. Thus, for those with more improvisation experience, function deviants should constitute a greater violation of expectation, and this should be predicted by greater discriminability of the N2c signal.

N2 components often occur in conjunction with P3 components. The P3 is elicited when participants detect relatively rare deviant stimuli in oddball tasks. It is also divided into subcomponents, namely the P3a and P3b (Polich, 2007). The P3a is elicited by the detection of deviants in oddball paradigms when a response is not required while the P3b, which is later in latency and more posterior topographically, occurs when a response is required by deviant detection. We thus expected our task to evoke the P3b. Stimulus probability modulates P3b activity, with less probable stimuli evoking stronger signals. However, it is not the objective probability of the stimuli; rather, how a participant categorizes stimuli can change the perceived *subjective* probability and thus also modulate the P3b (Johnson & Donchin, 1980). In our task, the exemplar and function deviants are close to equiprobable in occurrence, so the objective stimulus probability alone should not modulate the P3b activity, but those with more improvisation experience are hypothesized to categorize exemplars as similar to standards, making the function deviants *appear* relatively less probable, and relatively easier to detect. Because the subjective probability is lower, we expect those with more improvisation experience to show more discriminable P3b activity for function relative to exemplar deviants.

Overall, we predicted that behavioral differences in discriminability would be predicted by the discriminability of N2c and P3b activity. These components are thus analyzed as

**Table 1.** Participant demographic information.

Gender	24 M, 16 F
Age	$M = 25.28$ , $SD = 5.51$
Age of onset of musical training	$M = 11.18$ , $SD = 6.19$
Years of formal training on primary instrument	$M = 14.10$ , $SD = 7.21$
Number of participants who improvise < 1 hr/week (since age 18 and older)	15
Hours/week improvising (since age 18 and older) if $\geq 1$	$M = 10.38$ , $SD = 8.03$
Hours/week general playing but not improvising (since age 18 and older)	$M = 14.91$ , $SD = 12.44$
Piano players (primary instrument)	14
String players (primary instrument)	11
Wind players (primary instrument)	15

neurophysiological indexes of the information processing leading to participants' behavioral performance.

## Methods

### Participants

To take part in the study, participants had to have formal training (or significant professional experience) and play regularly, and those who had improvisation experience had to have formal training in improvisation. This allowed us to selectively include participants with a wide range of improvisatory ability. All participants reported having normal hearing, no history of neurological disorders, and proficiency on a pitched instrument. Refer to Table 1 for additional participant demographic information. The study took approximately 2 hrs and participants were remunerated for participating. Our university's Institutional Review Board approved all procedures and informed consent was obtained before the start of each experiment.

### Experimental design

Forty-one participants qualified and took part in the study. We excluded one participant's data for leaving the experiment before completion, leaving 40 participants ( $M = 25.28$  years old,  $SD = 5.52$ ; 24 male).

### Stimuli

Participants filled in a questionnaire with standard demographic information on age, gender, handedness—as assessed by the Edinburgh Handedness Inventory (EHI; Oldfield, 1971)—absolute pitch (self-reported yes or no), years of formal study on each instrument they played, years of those that included formal improvisation training, and self-identification as an improviser (on a scale from 1–7) and following Pinho et al. (2014), we also collected information on the number of hours per week participants played their primary instruments in the age ranges of under 11, between 12 and 17, and 18 and older, as well as the number of those hours that were specifically spent improvising. We instructed participants that this includes hours spent both practicing and performing and that they should be as accurate as possible. Few

participants asked for clarification about what counts as improvisation experience, but in those cases we encouraged participants to use their own judgment as to what constituted improvisation. For our final analysis, we decided only to examine how many hours per week participants played since the age of 18 because many participants started their training after the younger age range cutoffs and some could not accurately recall how much they played when they were younger. Further, we decided that this measure was a more accurate index of how active the musicians were at the time of the study. We thus obtained an index of improvisatory activity and experience (impHours), and an index of non-improvisatory activity and experience (nonImpHours), which was calculated by subtracting the impHours from the total hours indicated on our questionnaire.

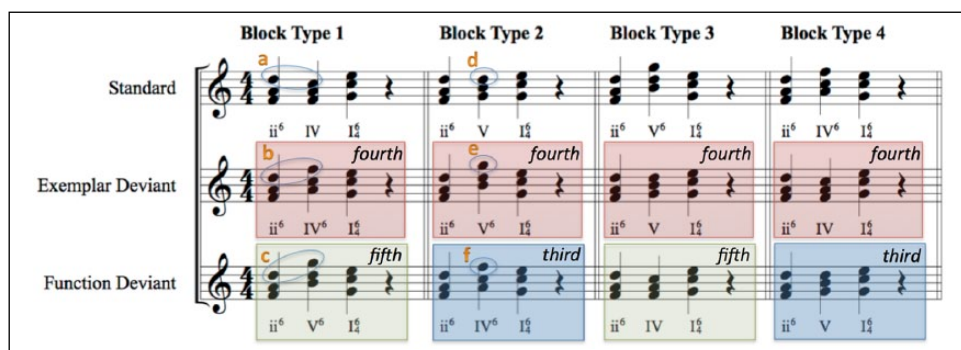
Because participants listed all instruments on which they have formal training, we were able to assess how many years of piano training each participant had. We included this exploratory variable, pianoYears, because we suspected that training specifically on the piano might affect participants' judgments on the auditory task owing to the piano timbre used for our stimuli.

Participants also briefly described their improvisation activities on the questionnaire. All participants were skilled in a Western tonal improvisatory style. Jazz was by far the most represented type of improvisatory music, although some participants also indicated experience with pop, classical, and folk improvisatory music in addition to jazz. We did not analyze our data with respect to this information, but further studies could examine differences between types of improvisatory experience (see Discussion).

We initially considered grouping participants into "improviser" and "non-improviser" categories on the basis of their self-identification as an improviser (4 or greater) and years of formal training (1 or more). However, we ultimately decided to use the impHours variable as an index of improvisatory experience because impHours provides a continuous measure, which allows us to show a continuous effect of the role of improvisation experience. The partial regression plots below retain the improviser/non-improviser grouping for visualization purposes.

The task required listening to different chord progressions of three chords each. The chord progressions were all either ii–IV–I, ii–V–I, ii–IV<sup>6</sup>–I, or ii–V<sup>6</sup>–I. A given chord progression could function as a standard or either type of deviant depending on the block type (see Figure 1). Each chord lasted 400 ms, with 400 ms of silence after the last chord (1600 ms total per trial). These 1600 ms stimuli were presented continuously, one after the other. There were 180 chord progressions per block, and eight blocks total in the experiment (two of each block type). There was also one practice block. Stimuli were created with a custom Python script using the pyMidi module to generate MIDI files, which were then converted to audio using a grand piano sound in GarageBand (Apple Inc., Cupertino, CA). The same stimuli were used for every participant, but the block order was randomized. A short break was taken after each block.

Each block's standard chord progression had a corresponding exemplar deviant and function deviant, each interspersed pseudorandomly with 7.5% probability (15% total deviants) such that a deviant was always followed by at least three standards, and each block always started with at least eight standard stimuli. Because deviants were dispersed probabilistically, the blocks did not have identical numbers of deviants (the combined set of stimuli from the four block types totaled 111 function deviants and 109 exemplar deviants). Deviant stimuli differed from the standard stimuli in the middle chord of the progression. In the exemplar deviant, the middle chord was substituted with a chord of the same musical function but a different inversion (e.g., a IV chord vs. a IV<sup>6</sup> chord in which the same notes are "stacked" in a different order). In the function deviant, the middle chord was substituted with a chord that was built on a different scale degree (e.g., a V chord instead of a IV chord), constituting a different functional class. Note that the middle chords of each kind of deviant have the same inversion. Further,



**Figure 1.** Stimuli used in the experiment. Each block type has three kinds of progressions. The different block types control for differences in contour (e.g., intervals a–c). However, note “e” is a fourth away from the expected standard note “d,” while note “f” is a third away from what would be expected in the standard. A larger deviation here could also potentially affect salience. According to this, stimuli can also be categorized into three groups (deviations of thirds, fourths, and fifths). We test the effect of this with a separate analysis that uses this grouping (see main text).

each chord progression was presented in a random key with the constraint that the same key was never presented twice in a row. The middle chord of each kind of deviant would thus be indistinguishable from each other in isolation; the type of deviant can only be determined by assessing the middle chord’s relationship to the first and third chords of the progression, i.e., by assessing its function. In improvisational practice, exemplar deviants would be much more likely to be substituted for the standard than the function deviants.

In a given block type, the three types of chord progressions also differ in terms of the intervals between the chords and the contour of the top notes. Contour could potentially change the relative salience of the two kinds of deviants (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004). We thus included four block types such that each chord progression could function as each condition (standard, exemplar deviant, and function deviant), allowing us to control for this variable and the number of common tones between chords (refer to Figure 1).

Finally, in any given block, the deviant conditions also differ from the standard in terms of the intervals of the top notes of the middle chord *relative* to the standard; this is similar to the previous point about intervals, but not the same. For example, if a participant listens to the top notes of the chords and expects a descending second but hears an ascending third, the deviation in the expectation of the interval (in this case, a fourth) might also affect the chord progression’s salience (Saarinen, Paavilainen, Schöger, Tervaniemi, & Näätänen, 1992). A larger interval difference could potentially be more salient in itself, regardless of functional categorization. Figure 1 illustrates this alternative categorization. We account for this with a separate analysis that compares the salience of each type of deviant progression based on these intervals rather than their grouping as function or exemplar deviants.

## Procedure

After completing the questionnaire, participants completed the oddball task while their EEG was recorded. Participants listened to the chord progressions over Bose in-ear noise cancelling headphones (Model QC-20) and were instructed to press the spacebar on a computer keyboard as quickly as possible if they heard a chord progression that differed “in any way” from what they perceived to be the most common chord progression within the context of that block. They

were told that the block would begin with several of the standard stimuli to help them judge what was the most common one local to that particular block (since the standard stimulus and corresponding types of deviants differed between blocks). They were also told to be accurate, i.e., to only press the spacebar if they are confident there is a difference. We collected reaction time and accuracy data in addition to EEG data. Importantly, participants were not told that there would be different types of deviant progressions, and were thus not encouraged to classify the stimuli in any particular way.

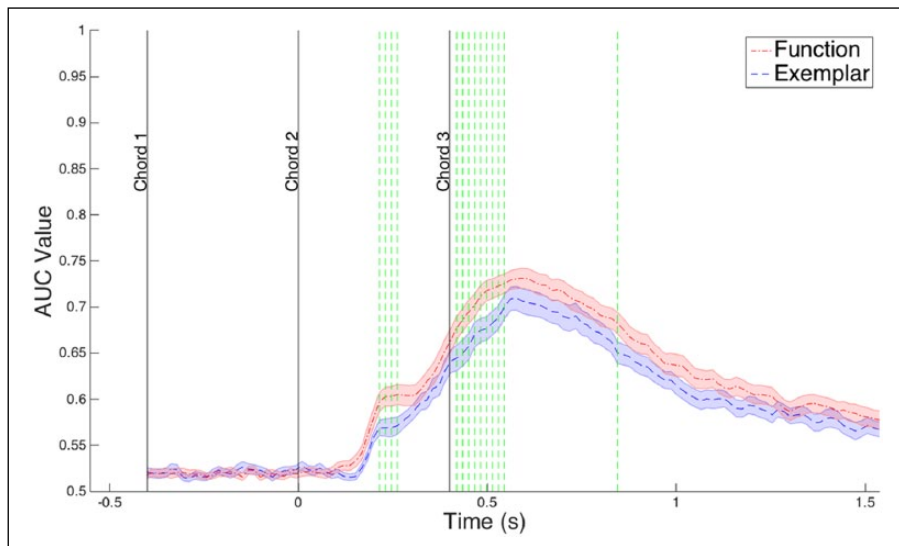
During the oddball task, EEG signal was acquired through 64 active scalp electrodes arranged according to the 10-20 system, sampled at 2048 Hz using a BioSemi Active Two AD Box ADC-12 amplifier (BioSemi, The Netherlands) in an electrostatically shielded room (ETS-Lindgren, Glendale Heights, IL, USA). The stimuli audio signal was also recorded on a separate channel to ensure proper alignment of the event markers, which were added offline after data collection. Separate computers were used to collect the EEG data and present the audio stimuli.

### ***Behavioral data analysis***

For each participant, we first computed the median reaction time of correct trials regardless of deviant type as well as the overall accuracy (total number correctly identified out of total number of deviants) of each participant. There was a statistically significant Pearson correlation between reaction time and accuracy values of the 40 participants,  $r_s(38) = -.39$ ,  $p = .012$ . Faster participants were also more accurate. Both of these values assess a participant's behavioral performance, so we captured information from both metrics by combining them into the Inverse Efficiency Score (IES), which is simply the reaction time divided by the accuracy (Townsend and Ashby, 1978, 1983). We then calculated the median RTs and accuracies for the individual deviant conditions separately and calculated an IES value for each that represented their behavioral performance for each deviant condition. Because the IES is a ratio variable, relatively small changes in low accuracy values can disproportionately affect the overall metric. To attenuate this effect, we log-transformed the IES values. Finally, for each participant, we subtracted the log-transformed IES value for function deviants from the value for exemplar deviants, yielding a single value, termed logIES, that we used as a dependent variable in statistical models described below. A higher logIES value means that the participant performed faster and more accurately for the function deviants relative to the exemplar deviants.

### ***EEG preprocessing***

EEG data were preprocessed using EEGLAB (Delorme & Makeig, 2004). Event markers for each chord were added offline using the audio channel data. Bad channels were interpolated using EEGLAB's inbuilt function, resulting in an average of  $3.16 \pm 0.25$  channels interpolated per block. Data were down-sampled from 2048 Hz to 64 Hz, band-pass filtered between 0.50 Hz and 30 Hz and re-referenced to the average activity of all electrodes. Eye movement artifacts were identified and rejected by eye using independent components analysis (ICA). Data from standard trials (not including false alarms) and deviant trials (only including hits) were epoched into 2000 ms epochs, ranging from 400 ms before the second chord of the progression to 1600 ms after it. In these epochs, 0 ms corresponds to the onset of the middle chord because that is the first point in time when a participant can possibly identify the deviants. The baseline was removed using the latency range from 100 ms before the onset of the middle chord. We excluded deviant epochs that had reaction times less than 200 ms from the onset of chord 2 as we considered these to be spurious. We used EEGLAB's functions to reject epochs based on extreme values in the EEG signal (outside of  $-250$  to  $250$   $\mu$ V) and based on probability (outside of 5 SD



**Figure 2.** AUC values calculated at sliding time windows over the epochs. A high AUC value means the sets of that type of deviant were more discriminable from standard epochs at that latency. The shaded area represents one standard error across participants in each direction. Green lines indicate points at which a paired *t*-test shows a significant difference between the curves ( $\alpha = 0.01$ ). Black lines show where the sound of each chord in the progression occurred. We did not consider the single window outside the two contiguous regions because it was temporally isolated and we thus assumed it reached significance by chance.

within channel, and across all channels). This left us with a set of exemplar deviant epochs ( $68.35 \pm 3.95$  per participant), function deviant epochs ( $78.25 \pm 3.06$  per participant) and standard deviant epochs ( $994.70 \pm 23.51$  per participant).

### *Linear discriminant analysis of EEG*

We used linear discriminant analysis (LDA) to determine the EEG activity that best discriminated each participant's set of exemplar deviants and function deviants from the standard epochs (correct trials only). For full details on how this is calculated and the advantages of this method over traditional ERP analysis, please see the Supplementary Materials online.

In short, the analysis provides a measure (the AUC value) of how discriminable each set of deviant epochs is from the set of standard epochs. AUC values are calculated over the course of the epoch showing the degree of discriminability at different latencies. Further, the analysis provides a forward model, which shows what combination of electrode activity best discriminates the sets of epochs at different times. This pattern of activity coupled with the latency information can be interpreted in similar terms as ERP analyses (Parra, Spence, Gerson, & Sajda, 2005).

## **Results**

### *Identifying regions of single-trial discriminability*

We conducted a two-tailed paired sample *t*-test ( $df = 39$ ) between the exemplar and function conditions for each point of the area under the curve (AUC) curves (see Figure 2). We used a



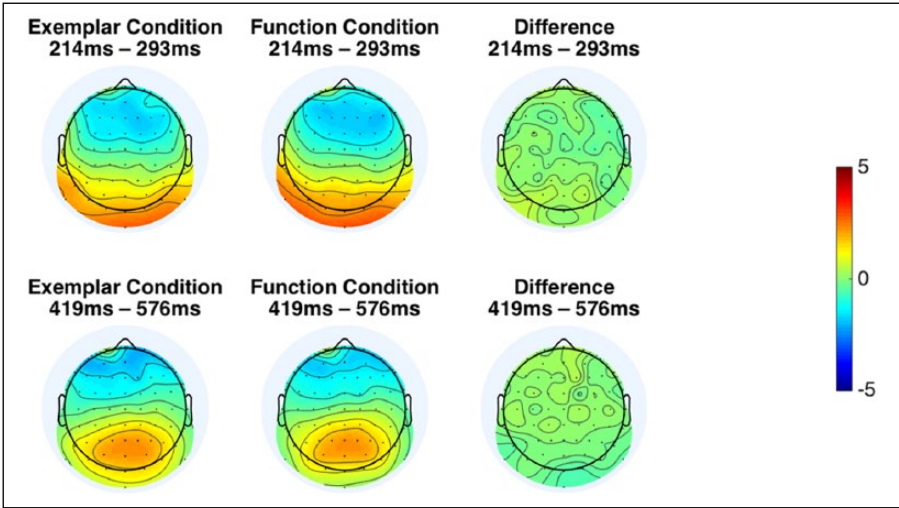
conservative alpha value of .01 for this test only (the others in our study use the standard .05 value). The  $t$ -tests revealed two fully contiguous regions where the curves differed significantly: an early range from 214 ms to 293 ms,  $t(39) = 3.08$ ,  $p = .004$ , on average in this region, and a later range from 419 ms to 576 ms,  $t(39) = 4.48$ ,  $p < .001$ , on average in this region. We considered this contiguity and occurrence in time windows of hypothetical interest (corresponding to N2 and P3 latencies) to be evidence against type I errors and thus did not further correct for multiple comparisons. We then computed the average forward model over the time points in these two time ranges (see Figure 3), which corresponds to the electrode activity that discriminates each type of deviant epoch from the standard epochs at those time ranges. The forward model corresponding to the earlier region has a frontocentral negativity that matches the latency and topography parameters of an auditory N2c ERP component, whereas the forward model for the later region has a posterior positivity that matches the latency and topography parameters of a P3b component. The topographies and latencies were very similar for function and exemplar deviants.

This difference can be interpreted as a main effect of condition—the function deviant has more discriminable N2c and P3b activity than the exemplar deviant—but to assess the relationship to the other variables of interest, for each participant, we averaged the AUC values in these two regions for each type of deviant and subtracted the exemplar value from the function value, yielding two values per participant that we termed EarlyAUC and LateAUC. A higher value for these variables means that the function deviants are more discriminable than the exemplar deviants on the basis of the corresponding forward model activity (N2c and P3b component activity respectively). We used these two values as dependent variables in our statistical modeling.

### *Behavioral and neural discriminability of deviant types*

We assessed the relationship between the three dependent variables logIES, EarlyAUC, and LateAUC, and the predictor variables impHours, nonImpHours, age, EHI score (from the handedness inventory), pianoYears, and the grouping variable absolute pitch (yes or no). We first conducted ANCOVAs to check for an effect of the grouping variable, absolute pitch, using type III sum of squares (because only eight participants had perfect pitch, leading to unequal group sizes). As this factor was not of experimental interest for our particular study, we wanted to ensure it was not interacting with the other continuous predictor variables in our models.

For the logIES and EarlyAUC variables, the ANCOVA showed no significant main effects or interactions with the absolute pitch categorical variable. However, in the LateAUC model, absolute pitch showed a significant interaction with nonImpHours,  $F(1, 28) = 4.45$ ,  $p = .044$ , and pianoYears,  $F(1, 28) = 6.03$ ,  $p < .020$ . This means that our covariates of interest affect people with and without absolute pitch differently. Absolute pitch has various effects on perceptual processing, and in particular has been linked to differences in P3 activity (Hantz, Crummer, Wayman, Walton, & Frisina, 1992), and, since we found that it was violating an assumption of ANCOVA modeling and was not a variable of hypothetical interest for the purposes of our study, we excluded perfect pitch participants ( $n = 8$ , 5 of whom were clustered into the non-improvisers group) from the LateAUC multiple regression models described below. For the sake of completion, we report the results of every multiple regression model both with and without the absolute pitch participants, although it is important to note that, strictly speaking, it is only necessary, statistically, to exclude these participants for the model in which there was a significant interaction in the ANCOVA (the LateAUC model).



**Figure 3.** Forward models corresponding to the two significant time regions identified in Figure 2. This can be interpreted as the electrode activity that discriminates each type of deviant from the standard chord progression activity. The exemplar and function conditions were not appreciably different in their topographies: in both, the early region corresponds to the latency and topography of the N2c ERP component and the later component corresponds to the latency and topography of the P3b component.

**Table 2.** Overall statistics for our multiple regression models.

Dependent variable	With absolute pitch participants						Without absolute pitch participants					
	Adjusted R <sup>2</sup>	df	df	F	p-value	Sig.	Adjusted R <sup>2</sup>	df	df	F	p-value	Sig.
logIES	.29	5	34	4.22	.004	***	.48	5	26	6.63	< .001	***
EarlyAUC	.01	5	34	0.93	.476		.04	5	26	1.23	.322	
LateAUC	.17	5	34	2.57	.045	*	.29	5	26	3.53	.014	*

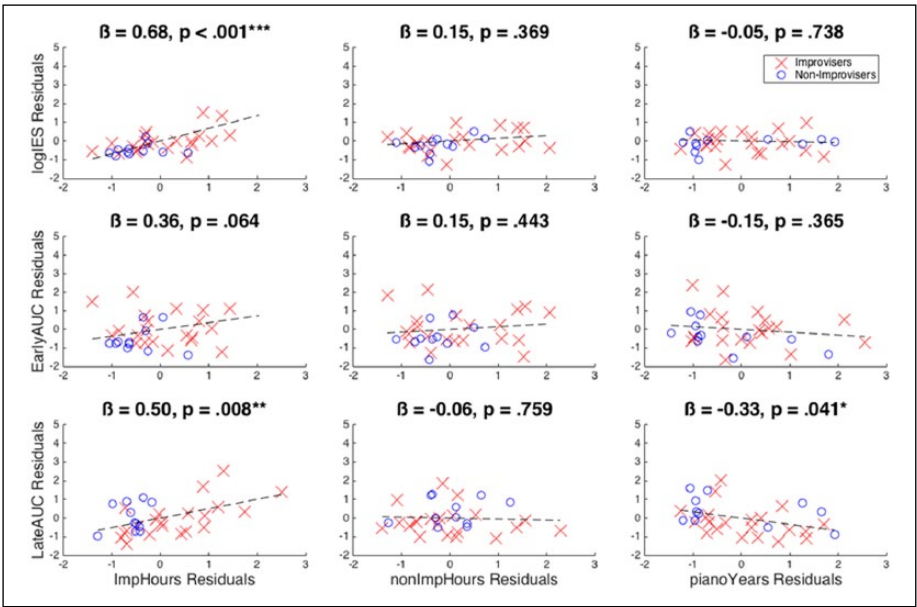
\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

We next performed a series of multiple regression models with the same independent and dependent variables as for the ANCOVAs, but without the grouping variable of absolute pitch. The overall statistics for the models are shown in Table 2, and the statistics for the individual regressors are shown in Table 3. The logIES model was significant overall, as was the LateAUC model. The impHours variable significantly predicted the logIES variable,  $\beta = 0.68$ ,  $p < .001$ . Those with more impHours performed relatively faster and more accurately for function deviants than for exemplar deviants. For the LateAUC model, after excluding participants with absolute pitch, impHours is a significant predictor,  $\beta = 0.50$ ,  $p = 0.008$ , showing more discriminability for function compared to exemplar deviants for those with more improvisation experience. Interestingly, pianoYears also significantly predicted the LateAUC variable,  $\beta = -0.33$ ,  $p = .041$ . Those with more piano training showed less discriminability between function and exemplar deviants in the LateAUC measure. The EarlyAUC model was not significant overall, although the effect of impHours itself was marginally significant,  $\beta = 0.36$ ,  $p = .064$ . The results of the multiple linear regressions are visualized in partial regression plots shown in Figure 4.

**Table 3.** Complete statistics with individual regressors for each our multiple regression models, both with and without the participants with absolute pitch included.

Dependent variable	Predictor variable	With absolute pitch participants			Without absolute pitch participants			
		Standardized $\beta$	$t$ -statistic	$p$ -value	Sig.	Standardized $\beta$	$t$ -statistic	$p$ -value
logIES	pianoYears	-0.05	-0.34	.738		-0.01	-0.09	.932
	nonImpHours	0.15	0.91	.369		0.07	0.45	.658
	impHours	0.68	4.27	<.001	***	0.79	5.27	<.001
	Age	0.04	0.25	.804		0.12	0.91	.370
	EHI Score	0.11	0.73	.469		0.04	0.32	.751
EarlyAUC	pianoYears	-0.15	-0.92	.365		-0.11	-0.61	.548
	nonImpHours	0.15	0.78	.443		0.22	1.06	.300
	impHours	0.36	1.92	.064	.	0.49	2.42	.023
	Age	-0.09	-0.51	.610		0.07	0.37	.714
	EHI Score	0.11	0.63	.536		0.03	0.18	.860
LateAUC	pianoYears	-0.43	-2.88	.007	**	-0.33	-2.15	.041
	nonImpHours	-0.20	-1.15	.257		-0.06	-0.31	.759
	impHours	0.18	1.06	.296		0.50	2.90	.008
	Age	0.06	0.37	.715		0.18	1.14	.267
	EHI Score	0.07	0.46	.648		0.004	0.03	.980

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .



**Figure 4.** Partial regression plots for the main analysis. Note that the logIES and EarlyAUC plots include participants with absolute pitch while the LateAUC plots do not due to the statistical violation detected by the ANCOVA (see main text). These plots show the partial contribution of each predictor variable to the overall linear models. Plots for EHI Score and age are not shown because neither of them showed significant effects and are not hypothetical variables of interest. All variables were standardized. Improviser and non-improviser group identity is distinguished for visualization only and was not included as a factor in our models. EHI: Edinburgh Handedness Inventory.

Finally, we tested the Pearson correlations between each of the dependent variables. logIES correlated with EarlyAUC,  $r(38) = .47$ ,  $p = .002$ , and with LateAUC,  $r(38) = .41$ ,  $p = .008$ , showing that behavioral performance correlates with N2c and P3b activity. EarlyAUC and LateAUC also significantly correlate with each other,  $r(38) = .36$ ,  $p = .021$ , indicating that greater N2c discriminability corresponds to greater P3b discriminability.

### Interval analysis

As mentioned above, stimuli can also be categorized according to the intervallic distance from the top note of the middle chord in each kind of deviant to the top note of the middle chord in the standard. We wanted to ensure the effects we found based on the function and exemplar stimuli categorizations were not being driven by this acoustic feature of the stimuli. We performed a similar analysis as with the main categorization of the stimuli. These models showed that larger intervals were *not* predicting greater discriminability in either the behavioral or neural data, preserving the validity of the findings from our main analysis. Please refer to the supplementary materials for more details of this alternative analysis.

### Discussion

In this study, we examined the effect of improvisational experience on the categorization of musical structures by testing how salient different kinds of deviants were in an oddball task. We hypothesized that those with more improvisational training would perceive exemplar deviants (chord progressions within a functional category) to be less salient relative to function deviants (chord progressions outside of a functional category) because of our theory that improvisers categorize musical structures based on function. This hypothesis was supported by both our behavioral and neural data. Those with more improvisation experience were better able to identify function deviants than exemplar deviants. Importantly, nonImpHours was not a predictive variable: it is specifically *improvisation* experience that predicts the differences.

The correlation between the behavioral measures to each of the neural measures lends consistency to our argument and allows us to tie in our explanation of this difference between these groups to the broader neuroscientific literature on categorization. Previous studies have found MMN and P3a differences between improvisers and classical musicians (Tervaniemi, Janhunen, Kruck, Putkinen & Huotilainen, 2015; Vuust et al., 2005, 2012). The N2c activity involved in other classification tasks and the way subjective probability modulates the P3b also evidently applies to our musical stimuli and can be leveraged to explain differences in how more experienced improvisers perceive and categorize stimuli: classification and categorization are indeed the relevant cognitive processes at play here.

More importantly, our study helps describe how improvisers perceive differently with specific regard to harmonic function. Thinking in terms of function may arise from improvisatory pedagogies and may aid the ability to improvise. Being able to “reach for” several possible realizations of functionally appropriate musical structures would aid the ability to improvise (cf. Sudnow, 1978). More experienced improvisers perceive different exemplars of a functional class as being similar, which provides evidence that their knowledge of structures is organized in this kind of way, and further research could explore how this is manifested in the neural instantiations of the representations of these structures.

Scientific observations of improvisers do not necessitate having participants actually improvise as they would outside the laboratory. Improvisation experience, as we have shown here, is associated with particular characteristics of how people perceive and categorize sounds.

Studying this perception and categorization is an indirect way to study improvisation abilities, but nevertheless advances an understanding of how people improvise, as is framed by our theoretical context outlined above. This link back to ecological practice should not be assumed, but we argue here that the research question is grounded well enough in an understanding of musical practice that in our case the link is strong.

Extrapolating our theories to other genres and traditions of improvisation is not trivial, but we believe it is possible. Improvisers in other traditions may also categorize different musical structures according to function (be it music-theoretical function or another kind of function), generalizing characteristics of specific exemplars within that category. Of course, to show how viable this is would require further theoretical inquiry and elaboration, and any experiments in other styles would rely on that foundation.

Explicit instructions can modulate deviant detection (Sussman, Winkler, Huottilainen, Ritter, & Näätänen, 2002), even in early perceptual processes indexed by the MMN. Our data suggest that participants did indeed categorize differently, but how explicit that categorization effort was remains unclear. Two interpretations with respect to our data are possible, and perhaps interrelated. First, those with more improvisation training may consciously choose to classify by function, which could worsen their detection of exemplar deviants. In debriefing interviews, few participants reported this, nor did they report noticing the different types of deviants. Further, we did not instruct participants to listen to any particular features of the stimuli to accomplish the task. Nevertheless, this would still need to be more conclusively ruled out. Another interpretation is that there is something different about improvisers' representations of musical structures that would predispose them to categorize this way, such that even if they tried to actively attend to other features, functional classes would still influence their perception. Future studies could specifically manipulate this variable by giving different categorization instructions to see if improvisers are still predisposed to hear function despite trying to attend to other characteristics of the stimuli.

A representational similarity analysis (RSA) study (Kriegeskorte, Mur, & Bandettini, 2008) could also build on our findings. This study assessed categorization on the basis of perceptual similarity. Given that our definition of musical function is tied to how a particular structure is *used*, the large and growing body of work on perception-action coupling in music (Maes, Leman, Palmer, & Wanderley, 2014; Novembre & Keller, 2014) and new promising work describing improvisation in terms of Gibsonian affordances (Love, 2017), it could be that motor areas are involved in representing such structures. A functional magnetic resonance imaging (fMRI) RSA analysis could help show what representational brain activity corresponds to the perceptual similarity. Functional classes of movements from other (non-musical) creative and everyday domains could also be assessed with a similar approach, similar to work on the neuroscience of goal representation (Hamilton & Grafton, 2006), and could be correlated with measures of creativity and improvisation ability.

Our findings show how certain organizations of knowledge may facilitate the ability to employ that knowledge in improvisatory behaviors. To be able to shift fluently between exemplars of a given functional class, as a similarity in representation would facilitate, could allow improvisers to have many different task-appropriate ways to behave in response to a particular stimulus. Note that this study situates knowledge organization as a relevant feature of improvisers' minds which bears on their ability to improvise; in this way, creative music performance is not addressed as a state of mind alone, but is predicated on particular ways of learning and knowing. Such a theory is cross-culturally compatible (cf. Nooshin, 2003) and compatible with theories of non-musical creative behaviors such as a chef substituting ingredients with a similar function or a martial artist responding to an attack with a variety of functionally similar blocks or parries. Tasks from the creativity literature like the Alternate Uses Test (cf. Arden,

Chavez, Grazioplene, & Jung, 2010) could also be understood in light of this theory. Speculatively, the ability to think of new uses of an object could be aided by organizing one's knowledge of objects by their function.

Differences in representational organizations could in principle be based on a kind of self-selection: those who have a predisposition to categorize by function for some other reason (be it inherent or learned in some other behavioral domain) may be more likely to become improvisers. Because our findings were based on a continuous predictor variable, the evidence from our study shows that the *amount* of experience at least contributes to the difference, highlighting the importance of training.

One alternative explanation for why knowledge organization differs comes from how improvising musicians are taught to value music structures. Students of jazz learn to make value judgments through a number of techniques including: reharmonizing pre-existing harmonic structures to actualize harmonic and melodic interaction, transcribing other musicians' solos or accompaniments to decode their "vocabulary" over harmonic changes, and comparing two or more pieces of music to find and utilize structural commonalities in their improvisational development (Berliner, 1994). These practices suggest, at least in part, that there is value placed on the utility of these music-structural elements, that is, how they assist the improviser in achieving the aural aesthetic that they desire. Taken this way, the functional deviant in our experiment could have presented as a structure with more utility for those with greater improvisation experience, thus pushing it to the front of the participants purview.

In contrast, Western non-improvising musicians may place value in completing challenging passages in a score at the desired tempo, or finding the right dynamics to accentuate a certain section of a composition. As such, educational practices in this environment emphasize playing sections of a composition and focusing on the dynamics, articulation, tone, and, expressivity (Zhukov, 2015), but not usually learning to manipulate functional-harmonic aspects of musical structures. As it relates to our experiment, a possible explanation for the non-improvisers' performance is that the non-improvisers placed value on more specific, local changes to how each chord in the 3-chord structure changed, such that exemplar and functional deviants were perceived as more equally different from the standard.

Finally, this work was motivated by the more general theory that the concept of improvisation can be understood as a particular way of knowing (Goldman, 2016). The structural differences in the knowledge of improvisers (e.g., differences in representational organization) are inextricable from the processes that they underlie in performance (i.e., the way they are able to employ them in action); because of this, understanding differences in the processes and the structures both contribute to an understanding of what it means to improvise. Difference in the categorization of structures on the basis of their function is one way improvisers have now been shown to differ (in addition to other ways they differ shown by the previous between-group comparison studies cited above). There may be many other ways that can meaningfully distinguish improvisatory ways of knowing.

## Acknowledgement

Sincere thanks to Josef Faller, Raphael Gerraty, Daphna Shohamy, George Lewis, Peter Gordon, Lori Custodero, the LIINC Lab, and the Learning Lab for supporting this project. Thanks also to the Presidential Scholars in Society and Neuroscience program.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclose receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by the Presidential Scholars in Society and Neuroscience program at Columbia University, and was partially supported by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-15-2-0074.

## References

- Arden, R., Chavez, R. S., Grazioplene, R., & Jung, R. E. (2010). Neuroimaging creativity: A psychometric view. *Behavioural Brain Research*, 214, 143–156.
- Bengtsson, S. L., Csikszentmihályi, M., & Ullén, F. (2007). Cortical regions involved in the generation of musical structures during improvisation in pianists. *Journal of Cognitive Neuroscience*, 19, 830–842.
- Berkowitz, A. L., & Ansari, D. (2008). Generation of novel motor sequences: The neural correlates of musical improvisation. *Neuroimage*, 41(2), 535–543. doi:10.1016/j.neuroimage.2008.02.028
- Berkowitz, A. L., & Ansari, D. (2010). Expertise-related deactivation of the right temporoparietal junction during musical improvisation. *Neuroimage*, 49(1), 712–719. doi:10.1016/j.neuroimage.2009.08.042
- Berliner, P. (1994). *Thinking in jazz: The infinite art of improvisation*. Chicago: University of Chicago Press.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21.
- de Manzano, Ö., & Ullén, F. (2012). Goal-independent mechanisms for free response generation: Creative and pseudo-random performance share neural substrates. *NeuroImage*, 59, 772–780.
- Donnay, G. F., Rankin, S. K., Lopez-Gonzalez, M., Jiradejvong, P., & Limb, C. J. (2014). Neural substrates of interactive musical improvisation: An fMRI study of 'trading fours' in jazz. *PLoS ONE*, 9(2), 1–10.
- Frieler, K., Pfeleiderer, M., Zaddach, W. -G., & Abeßer, J. (2016). Midlevel analysis of monophonic jazz solos: A new approach to the study of improvisation. *Musicae Scientiae*, 20, 143–162.
- Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience*, 16, 1010–1021.
- Goehr, L. (1992). *The imaginary museum of musical works*. Oxford: Oxford University Press.
- Goldman, A. (2013). Towards a cognitive-scientific research program for improvisation: Theory and an experiment. *Psychomusicology*, 23, 210–221.
- Goldman, A. (2016). Improvisation as a way of knowing. *Music Theory Online*, 22(4). Retrieved from: <http://mtosmt.org/issues/mto.16.22.4/mto.16.22.4.goldman.html>
- Hamilton, A. F. d. C., & Grafton, S. T. (2006). Goal representation in human anterior intraparietal sulcus. *Journal of Neuroscience*, 26, 1133–1137.
- Hantz, E. C., Crummer, G. C., Wayman, J. W., Walton, J. P., & Frisina, R. D. (1992). Effects of musical training and absolute pitch on the neural processing of melodic intervals: A P3 event-related potential study. *Music Perception*, 10, 25–42.
- Hargreaves, W. (2012). Generating ideas in jazz improvisation: Where theory meets practice. *International Journal of Music Education*, 30, 354–367.
- Harris, R., & de Jong, B. M. (2015). Differential parietal and temporal contributions to music perception in improvising and score-dependent musicians, an fMRI study. *Brain Research*, 1624, 253–264.
- Johnson, R., & Donchin, E. (1980). P300 and stimulus categorization: Two plus one is not so different from one plus one. *Psychophysiology*, 17, 167–178.
- Johnson-Laird, P. N. (1991). Jazz improvisation: A theory at the computational level. In P. Howell, R. West & I. Cross (Eds.), *Representing musical structure* (pp. 291–325). London: Academic.
- Johnson-Laird, P. N. (2002). How jazz musicians improvise. *Music Perception*, 19, 415–442.
- Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2, 1–28.
- Lewis, G. E., & Piekut, B. (Eds.). (2016). *The Oxford handbook of critical improvisation studies*. New York: Oxford University Press.
- Limb, C. J., & Braun, A. R. (2008). Neural substrates of spontaneous musical performance: An fMRI study of jazz improvisation. *PLoS ONE*, 3(2), e1679.

- Liu, S., Chow, H. M., Xu, Y., Erkkinen, M. G., Swett, K. E., Eagle, M. W., . . . Braun, A. R. (2012). Neural correlates of lyrical improvisation: An fMRI study of freestyle rap. *Scientific Reports*, 2, 834. doi:10.1038/srep00834
- Love, S. C. (2017). An ecological description of jazz improvisation. *Psychomusicology*, 27, 31–44. doi:10.1037/pmu0000173
- Maes, P.-J., Leman, M., Palmer, C., & Wanderley, M. M. (2014). Action-based effects on music perception. *Frontiers in Psychology*, 4, 1008. doi: 10.3389/fpsyg.2013.01008
- Nooshin, L. (2003). Improvisation as 'other': Creativity, knowledge and power—The case of Iranian classical music. *Journal of the Royal Musical Association*, 128, 242–296.
- Norgaard, M. (2011). Descriptions of improvisational thinking by artist-level jazz musicians. *Journal of Research in Music Education*, 59, 109–127.
- Norgaard, M. (2014). How jazz musicians improvise: The central role of auditory and motor patterns. *Music Perception*, 31, 271–287.
- Norgaard, M., Spencer, J., & Montiel, M. (2013). Testing cognitive theories by creating a pattern-based probabilistic algorithm for melody and rhythm in jazz improvisation. *Psychomusicology*, 23, 271–287.
- Novembre, G., & Keller, P. E. (2014). A conceptual review on action-perception coupling in the musicians' brain: What is it good for? *Frontiers in Human Neuroscience*, 8, 1–11.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97–113.
- Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P. (2005). Recipes for the linear analysis of EEG. *Neuroimage*, 28, 326–341.
- Pfleiderer, M., & Frieler, K. (2010). The Jazzomat project. Issues and methods for the automatic analysis of jazz improvisations. In R. Bader, C. Neuhaus & U. Morgenstern (Eds.), *Concepts, experiments, and fieldwork: Studies in systematic musicology and ethnomusicology*. (pp. 279–295). Frankfurt: Peter Lang.
- Pinho, A. L., de Manzano, Ö., Fransson, P., Eriksson, H., & Ullén, F. (2014). Connecting to create: Expertise in musical improvisation is associated with increased functional connectivity between premotor and prefrontal areas. *Journal of Neuroscience*, 34, 6156–6163.
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118, 2128–2148.
- Pressing, J. (1988). Improvisation: Methods and models. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 129–178). Oxford: Clarendon Press.
- Pressing, J. (1998). Psychological constraints on improvisational expertise and communication. In B. Nettl & M. Russell (Eds.), *In the course of performance: Studies in the world of musical improvisation* (pp. 47–67). Chicago and London: University of Chicago Press.
- Ritter, W., Simson, R., & Vaughan, H. G. (1983). Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks. *Psychophysiology*, 20(2), 168–179.
- Saarienen, J., Paavilainen, P., Schöger, E., Tervaniemi, M., & Näätänen, R. (1992). Representation of abstract attributes of auditory stimuli in the human brain. *NeuroReport*, 3, 1149–1151.
- Sudnow, D. (1978). *Ways of the hand: The organization of improvised conduct*. Cambridge: Harvard University Press.
- Sussman, E., Winkler, I., Huottilainen, M., Ritter, W., & Näätänen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. *Cognitive Brain Research*, 13, 393–405.
- Tervaniemi, M., Janhunen, L., Kruck, S., Putkinen, V., & Huottilainen, M. (2015). Auditory profiles of classical, jazz, and rock musicians: Genre-specific sensitivity to musical sound features. *Frontiers in Psychology*, 6, 1–11.
- Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In J. Castellan & F. Restle (Eds.), *Cognitive theory* (pp. 200–239). Hillsdale, N.J.: Erlbaum.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.



- Vuust, P., Brattico, E., Seppänen, M., Näätänen, R., & Tervaniemi, M. (2012). The sound of music: Differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. *Neuropsychologia*, 50, 1432–1443.
- Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstorff, A., & Østergaard, L. (2005). To musicians, the message is in the meter: Pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *NeuroImage*, 24, 560–564.
- Zhukov, K. (2015). Challenging approaches to assessment of instrumental learning. In D. Lebler, G. Carey & S. D. Harrison (Eds.), *Assessment in music education: From policy to practice* (pp. 55–70). Cham: Springer International.