Quantifying the Benefits of Using an Interactive Decision Support Tool for Creating Musical Accompaniment in a Particular Style

Ching-Hua Chuan
University of North Florida
School of Computing
Jacksonville, FL
chchuan@mail.barry.edu

Elaine Chew
University of Southern California
Viterbi School of Engineering
Thornton School of Music
Los Angeles, CA
echew@usc.edu

ABSTRACT

We present a human-centered experiment designed to measure the degree of support for creating musical accompaniment provided by an interactive composition decision-support system. We create an interactive system with visual and audio cues to assist users in the choosing of chords to craft an accompaniment in a desired style. We propose general measures for objectively evaluating the effectiveness and usability of such systems. We use melodies of existing songs by Radiohead as tests. Quantitative measures of musical distance – percentage correct and closely related chords, and average neo-Riemannian distance – compare the user-created accompaniment with the original, with and without decision support. Numbers of backward edits, unique chords explored, and repeated chord choices during composition help quantify composition behavior. We present experimental data from musicians and non-musicians. We observe that decision support reduces the time spent in composition, the number of revisions of earlier choices, redundant behavior such as repeated chord choices, and the gap between musicians' and non-musicians' work, without significantly limiting the range of users' choices.

1. INTRODUCTION

Building computer-assisted composition software tools has long been a common interest among researchers. Besides tools designed for professional use with advanced functions such as audio signal editing, many aim to help amateurs and music lovers write music and to enhance users' musical creativity. These tools often provide alternate ways for representing music, using interfaces such as a drawing pad to avoid score notation. While various systems have been proposed in recent years, it remains unclear how much these systems help the composition process. Evaluations of systems that support art creation often take the form of subjective opinion with varying degrees of success, and cannot be used to systematically study and improve the methodology. Music compositions, like other art forms, are often evaluated using the Consensual Assessment Technique [10], which measures quality according to experts' global and subjective assessment of the outcome. For studies that aim to design tools to support music composition, this technique may not provide sufficient detail for evaluating the systems. For example, it would be difficult to determine to what extent users' creations are improved by a composition assistance system using the Consensual Assessment Technique.

In this paper we focus on experiments that aim to quantifying the added benefit of style-imitating accompaniment composition decision-support systems. Such systems aim to help users create accompaniment to a melody in the style of a particular artist of band. When the goal is to imitate the style of a known song, the original accompaniment serves as ground truth for any user-created accompaniment to the song's melody. If the user is able to produce more chord patterns that are similar to the original accompaniment with automated assistance, it can then be concluded that the system achieves its design goal.

For the experiments described in this paper, we built a composition decision-support software tool by adapting an automated system for generating style-specific accompaniment. We designed an intuitive interactive user interface that assumes no formal music knowledge, using visual cues and sound to make music composition feasible for experts and novices alike. The system provides suggestions to the user at two levels: the system generates a sequence of stylistically appropriate chords as initial suggestions, a range of triads is then listed to guide users in each chord selection.

To measure the benefits of the system, we conducted a human-centered experiment with three composition tasks. In each task, participants were given a melody and were requested to complete the composition by choosing chords to accompany the melody in the style of a particular artist. Before the experiment, the participants listen to several songs by the artist they should emulate. In the first task, participants were given a composing interface without decision support; in the latter two tasks, they worked with a composing interface with decision-support that put forward sequences of computer-suggested chords. The experiment was designed to explore the composition process by examining whether participants could mimic an artist's composition decisions after having heard samples of their songs, and by determining how much their performance improved when they worked with a composition decision-support system.

We separated the participants into two groups – musicians and non-musicians – according to their musical backgrounds, and evaluated the effectiveness of the composition decision-support tool by examining whether non-musicians, with the help of the system, can create accompaniments that are as close to the original style as those by musicians. We propose quantitative metrics for evaluating the system's usability by statistically examining the changes in the users' composition behaviors with and without decision support.

2. RELATED WORK

In addition to the systems that are designed to model music composition [1, 4, 13, 15], efforts have been made to build tools to support music composition [2, 7, 8]. Unlike music production software that only provide tools for notating scores and recording sound, software such as Hyperscore [8] and Sonic Sketchpad [7] offer alternative modalities (e.g., drawing) through which to compose music. With this kind of software, users do not need to be familiar with, or knowledgeable about, music notation or instrument-specific characteristics. Evaluations of these systems focus on users' level of satisfaction with the software interface and its functionalities.

In particular, the automatic style-specific accompaniment (ASSA) system described in [5] aims to assist amateurs in songwriting by helping them compose music in the style they desire. The system incorporates both music theoretic knowledge and statistical learning, using small numbers of examples of users' favorite artists to model the accompaniment choices of that professional. It then applies the learned composition style to new, user-created melodies. The ASSA system has been evaluated based on subjective opinion through a Turing test, and on style-related quantitative metrics [6].

3. SYSTEM DESIGN

3.1 Interactive User Interfaces

In this study, the ASSA system [5] forms the basis of our composition decision-support tool. The interactive user interface allows users to create accompaniment to any melody through simple mouse-click actions on graphic icons. Users can explore different chords in each bar and listen to parts of the melody in any order with the accompaniment they create. Users can also adjust the tempo of the song to better fit their listening pace during composition. The system takes care of composition details such as voice arrangement and chord alignment, leaving users free to concentrate on using their aural perception to select the appropriate chords.

To test the improvement resulting from computer assistance in music composition, we created two interface designs. The first interface, shown in Figure 1, provides users with all possible chord choices at all times, i.e. without decision support. The chord collection, all 24 triads (major and minor), appears in the top panel of the interface. Triads are arranged according to the circle-of-fifths, with the major chord cycle on the left and the minor triad cycle relative to the major on the right. The

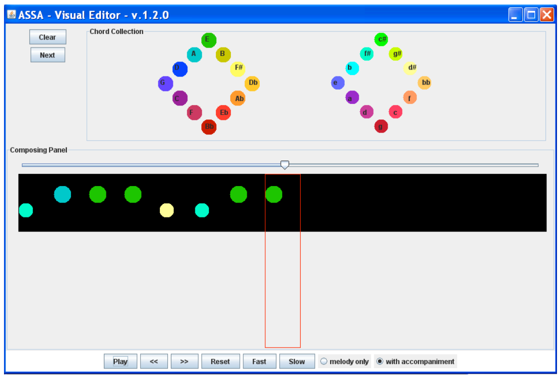


Figure 1. Composing interface without ASSA support.

Figure 2. Composing interface with ASSA support and graph of neo-Riemannian distance.

tonic triad is centered at the top. Each triad is assigned a different color to help users identify chords and the relationships between them. We used the color assignment described in [12], in which related colors are assigned to chords considered close one to another.

During composition, users can choose chords by clicking on one of the colored circles in the chord collection. Once a chord is selected, a corresponding triad with proper pitch arrangements is inserted in the accompaniment MIDI track, and a circle of the same color is placed in the composing panel, the black rectangle in the center of Figure 1. The composing panel allows users to visually examine the chord sequence they create. Major triads appear as larger colored disks, minor triads as smaller ones.

Figure 2 depicts the second interface. The most visible difference between this interface and the previous one is the graph at the bottom of the interface. Each point on the grid represents the neo-Riemannian distance [3] between adjacent chords. In neo-Riemannian chord space, chords are connected by neo-Riemannian operations (NROs), and the number of operations between chords reflects their musical distance. The user can select the time slice to be examined, and edit the neo-Riemannian distance they wish to consider at that time slice. Other chords within that number of NROs from the previous chord are displayed in the top panel. In this manner, the number of chord choices is reduced and constrained by the NRO distance, which helps users categorize chords into subgroups that reflect their musical distance in the transition.

In addition to the use of neo-Riemannian distance, the user starts with an ASSA-produced chord sequence and its corresponding sequence of neo-Riemannian distances. The providing of this initial suggestion simplifies the process of creating accompaniments from scratch to one of editing the chords until the user is satisfied with the accompaniment.

4. EXPERIMENT DESIGN

Our study aims to determine the effectiveness of a composition decision-support system through quantitative comparisons of users' compositions, and of their decision pathways, with and without decision support. This section presents our experiments to further these goals.

4.1 Experiment Setup

We designed a three-part experiment using the two composition interfaces described in the previous section. In the first part of the experiment, participants were given a melody extracted from a song with which they are familiar, and were asked to compose a sequence of chords without decision support, using the interface shown in Figure 1. In parts two and three, participants were presented with the interface shown in Figure 2, and asked to compose accompaniments for two unfamiliar melodies in a familiar style with decision support.

We used songs by Radiohead, a British rock band, for the experiment. We chose Radiohead not only because of the band's popularity and reputation, but also because their unique style has been analyzed by music theorists in several book-length studies [9, 11, 14]. We selected 13 songs from Radiohead's albums *Pablo Honey* and *The Bends* as samples to familiarize participants with the band's style and as the training set for the ASSA system. As test melodies for the three-part experiment, we extract the melodies of three "hit" songs from the two albums, as indicated by their popularity ratings on the iTunes online music store. The participants' training samples included the test song for the first part, but not for the second and third parts. Each participant was asked to complete all three parts in the experiment. The idea behind these choices was to investigate whether participants, especially non-musicians, could create accompaniments for new melodies (the second and third tasks) in a style similar to the original with decision support.

4.2 Experiment Procedure

The participants for this study consisted of 26 volunteers: 8 musicians and 18 music lovers. The majority of the non-musician participants were college students from various disciplines. Participants who were categorized as musicians either majored in Music at a university or played an instrument in a rock band. Participants were requested to spend at least three days listening to the sample Radiohead songs provided, and to schedule an appointment for the experiment after they are familiar with the band's style.

The experiment was conducted in a computer lab, where each participant was assigned a computer and a headset to complete the task individually. Before beginning, participants first watched a 10-minute instructional video on how to use the two interfaces to create accompaniments. Before they started composing with the software, participants were reminded to choose chords that they think Radiohead would choose, based on their experience listening to music by the band. Participants were then left alone to compose without time constraints.

4.3 Evaluation Metric

This section describes the evaluation criteria designed to fulfill the aforementioned goals.

4.3.1 Effectiveness

The effectiveness of system assistance can be measured in many ways. If musicians as well as non-musicians take less time to complete their compositions with the decision support, we can say that the system helps users by speeding up the composition process.

Decision support effectiveness can also be measured in the improvement of musicians' and non-musicians' compositions produced with (vs. without) decision support, and of non-musicians' compositions with regard to their similarity to musicians' compositions. For instance, if non-musicians create accompaniments more similar to the ones by musicians with decision support than without, it can be claimed that the decision support helps non-musicians make more knowledgeable decisions to better reach the targeted music style.

To measure the quality of the user-created accompaniments, we compared them to that of the original production, as documented in the commercial sheet music. We counted the number of chords in the user-created accompaniment that are identical to the original, and divided it by the total number of chords to report the *same chord* percentage. For the chords in the user-created accompaniment that are different from the original, we further determined the percentage of chords that are closely related to the original. The metric, *chords-in-grid,* reports the percentage of chords that are related to the original by a Dominant (D), Subdominant (S), Parallel (P), Relative (R), or by one of the compound relations: DP, SP, DR, and SR. We also measured the *average NR distance* between the user-created and the original accompaniment to assess how dissimilar the user-chosen chords are from the original in terms of NROs.

4.3.2 Usability

Another goal of the study is to examine how the decision support system affects users' composition processes. We analyzed the changes in the way chords are selected (without decision support) or edited (with decision support) to investigate whether the decision support guides users in their compositions or limits their choices. Specifically, we focused on three types of behavior patterns: *backward editing*, *exploration*, and *hesitation*.

Backward editing occurs when users change the chord in a bar that is before the current one after they select a chord in the current position. For example, the following action sequence contains two backward edits at times T_2 and T_4, respectively:

```
Bar:              5    4    5    3
Chord selected:   C    Am   G    Dm
Time:             T1   T2   T3   T4
```

Backward editing indicates that the user has become uncertain about a previous choice they made because of a later decision.

We define *exploration* as the average number of unique chords that users have tried assigning in each bar, including the ones they choose and then changed later. For example, in the scenario above, two unique chords were explored in bar 5 at time T_1 and T_3 respectively. In contrast, *hesitation* is defined as the total number of repeated chord choices during the entire composition process. For instance, the following sequence contains a repeated choice at time T_4 in bar 6:

```
Bar:              5    6    6    6
Chord selected:   C    Am   C    Am
Time:             T1   T2   T3   T4
```

For each of the two groups, musicians and non-musicians, we calculated the means and standard deviations of the *backward editing*, *exploration*, and *hesitation* values with and without decision support. Suppose $\mu_{1,m,b}$ represents the true mean of musicians' backward editing without decision support and $\mu_{2,m,b}$ the true mean with decision support. We conducted a hypothesis test on whether decision support has a positive impact on the behavior pattern (i.e. reduces backward editing) by setting the null hypothesis, H_0, and the alternative hypothesis, H_1, as:

$$H_0: \mu_{1,m,b} \leq \mu_{2,m,b}$$
$$H_1: \mu_{1,m,b} > \mu_{2,m,b} \qquad (1)$$

We calculated the observed level of significance, the p-value, using the Student's t-distribution with the calculated means and standard deviation. We chose the Student's t-distribution over the normal distribution due to the small sample size. We then compare the p-value with the standard 5% significance level. Similar hypothesis tests were conducted as well for the exploration and hesitation mean values.

5. EXPERIMENT RESULTS AND ANALYSES

In this section, we report the results of the experiments, and describe the analyses of these results. Twenty-six participants volunteered for the experiment, each spending approximately about one hour in completing the tasks. These users, and their composition results, are separated into two groups, namely musicians and non-musicians, according to their musical backgrounds.

5.1 Effectiveness: Inter-Group Comparisons

This section presents analyses of composition decision support effectiveness within the musician and non-musician groups.

5.1.1 Task Completion Time

Table 1 lists the average time musicians and non-musicians spent on each composition task. For song 1, which was included in the participants' training samples, musicians and non-musicians spent an almost equal amount of time choosing the 15 chords for the accompaniment without decision support. For songs 2 and 3, which were new to the participants, non-musicians spent a slightly shorter amount of time generating an accompaniment with an initial computer-produced suggestion followed by edits guided by the neo-Riemannian interface. If we calculate the time spent on each chord in the accompaniment, we can observe that both musicians and non-musicians spent significantly less time on the second and third task than on the first.

Time (min)	**Song 1 (15 chords)**	**Song 2 (36 chords)**	**Song 3 (30 chords)**
Musicians	17.54	20.59	12.76
Non-musicians	17.51	14.13	7.77

Table 1. Time spent on each accompaniment task.

5.1.2 Created Accompaniment versus the Original

Figure 4 shows the mean *same chord*, *chords in grid*, and *average NR distance* percentages for the user-created accompaniments. The results of accompaniments produced with decision support are the average of that for song 2 and song 3.

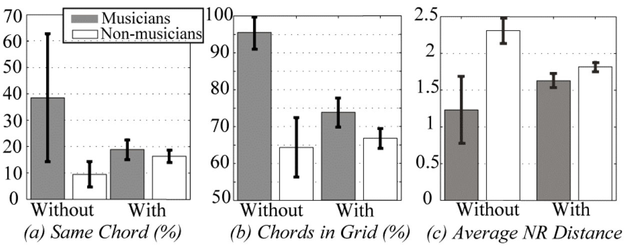


Figure 4. Evaluation of user-created accompaniments.

Figure 4(a) shows that the average percentage of chords in the user-created accompaniment that are identical to those in the original accompaniment is much higher in compositions by musicians than in those by non-musicians. This performance gap between musicians and non-musicians is greatly reduced with the assistance of the ASSA system. Similar patterns can be observed in Figure 4(b) with the *chords-in-grid* metric and in Figure 4(c) with the *average NR distance* metric.

If we focus on the musicians' results, we will find that, even with decision support, musicians performed worse on songs 2 and 3, when they are unfamiliar with the test song, than on song 1, when they had heard the song in its entirety. This finding implies that it is more difficult for musicians to come up with exactly the same chords as the original if they have never heard the song before. In contrast, non-musicians chose more suitable chords for songs 2 and 3, in collaboration with the system, than they did for song 1. With decision support, as in songs 2 and 3, the performance of non-musicians approached that of musicians.

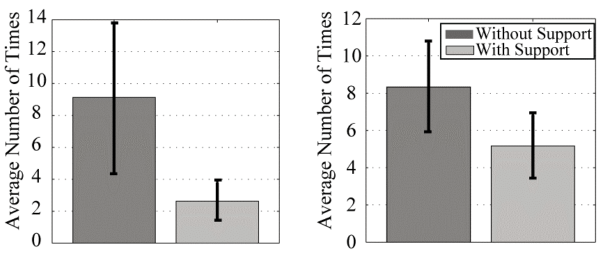
Note that in the bars labeled *without* in Figures 4(a) and 4(b), the standard deviation for the percentage of *same chords* for the musicians' accompaniments is much greater than that for *chords-in-grid* for the same group. This difference is due to the fact that the musician group consists of classical music majors as well as rock band members. It appears that imitating Radiohead's style is a challenge for some classical musicians who were less familiar with the band than rock musicians. However, because of their training in music theory, their choices are closely related to the original as the high *chords-in-grid* values suggest.

5.2 Usability: Intra-Group Comparisons

This section presents analyses of user composition patterns, with and without decision support, comparing the results of musicians against that of non-musicians.

5.2.1 Backward Editing

Figure 5 presents the average number *backward edits* with and without decision support for (a) musicians and (b) non-musicians. Note that for both musicians and non-musicians, decision support reduces the amount of *backward editing*. The results are confirmed statistically by the p-values; both the p-values for musicians (p-value = 0.0066) and non-musicians (p-value = 0.015) are less than 0.05, indicating that the reduction in backward edits is statistically significant.

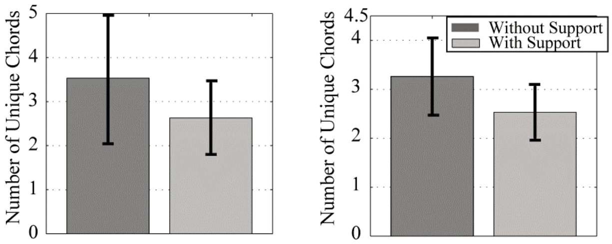


(a) Musicians (p-value = 0.0066) *(b) Non-Musicians (p-value = 0.015)*

Figure 5. Average backward edit counts with and without the decision support.

5.2.2 Exploration

Figure 6 shows the average number of unique chords explored in each bar for (a) musicians and (b) non-musicians, with and without the decision support.



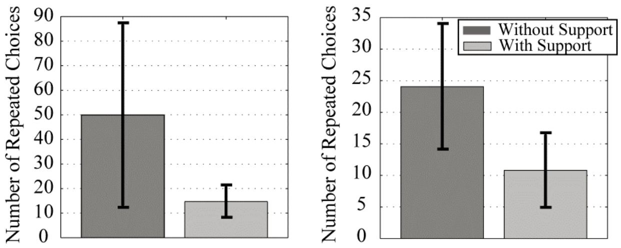
(a) Musicians (p-value = 0.112) *(b) Non-Musicians (p-value = 0.057)*

Figure 6. Average number of unique choices explored in each bar with and with the decision support.

While both musicians and non-musicians explore fewer numbers of unique chords on average with system support, the p-values are higher than 0.05. Thus, the observed reduction in unique chords explored is not statistically significant, and we conclude that the system does not significantly limit participants' exploration.

5.2.3 Hesitation

The average number of repeated choices in each bar with and without decision support are shown in Figure 7 for (a) musicians and (b) non-musicians. It can be observed that the number of repeats is reduced when the system provides suggestions during composition, as reflected in the figures as well as the p-values. The result implies that the system helps musicians and non-musicians reach their goals with less confusion.



(a) Musicians (p-value = 0.03) *(b) Non-Musicians (p-value = 0.015)*

Figure 7. Average number of repeated choices in each bar with and without decision support.

6. CONCLUSION AND FUTURE WORK

In this paper we described an interactive decision support system that aims to assist amateurs in the creating of accompaniments in a desired style. The system uses visual cues to offer chord suggestions and sounds to evaluate chord choices, making composition feasible for people without formal musical training. To investigate the benefits associated with the decision support, we designed three composition tasks. In the first task, participants were provided with the interface without decision support, and asked to create an accompaniment for a melody of a familiar song. In the latter two tasks, participants were given the interface with decision support, and instructed to produce accompaniments for unfamiliar melodies in a familiar style.

With the data obtained in the experiment, we first analyzed how well the user-created accompaniments concurred with the desired style, separating the results by user group, using the quantitative measures: same chord and closely-related chord percentages, and average neo-Riemannian distance. We observed that when decision support was present, the performance gap between musicians and non-musicians was greatly reduced.

To understand the usability of the system, we proposed methods to measure the changes in participants' composition behavior. Statistical analyses revealed that when decision support was available, the number of backward edits and repeat choices were reduced, and the range of unique chords explored was not significantly limited by the system.

Additional analyses that can be performed on the data include examining whether common or distinct decision patterns exist between the two user groups, and separating the support given by the initial ASSA-produced sequence from the assistance provided by the interactive interface and differentiating between them.

In the future, we plan to extend the experiment by considering other musical factors such as timbre and instrument arrangement to investigate the impact of different stimuli on musical creativity. We will refine the interface and conduct studies with larger groups of users, and design questionnaires to obtain their feedback on the system. Last but not least, we will further study the degree to which automation affects music creativity so as to design better composition decision support tools.

7. ACKNOWLEDGEMENTS

We thank the volunteers who participated in the experiment, including students and alumni of the University of Southern California and of Barry University. This work was supported in part by NSF Grant No. 0347988. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors, and do not necessarily reflect those of the NSF.

8. REFERENCES

[1] M. Allan and C. K. I. Williams: "Harmonising Chorales by probabilistic inference." *Proceedings of the Neural Information Processing Systems Conference*, Vancouver, pp. 25–32, 2004.

[2] J. Bamberger: "Developing Musical Intuitions: A Project-Based Introduction to Making and Understanding Music." Oxford University Press, USA.

[3] G. Capuzzo: "Neo-Riemannian theory and the analysis of pop-rock music." *Music Theory Spectrum*, Vol. 26, No. 2, pp. 177–199, 2004.

[4] H. Chan and D. Ventura: "Automatic Composition of Themed Mood Pieces." *Proceedings of the 5th International Joint Workshop on Computational Creativity*, Madrid, pp. 109–115, 2008.

[5] C. H. Chuan and E. Chew: "A Hybrid System for Automatic Generation of Style Specific Accompaniment." *Proceedings of the 4th International Joint Workshop on Computational Creativity*, London, pp. 57–64, 2007.

[6] C. H. Chuan and E. Chew: "Evaluating and Visualizing Effectiveness of Style Emulation in Musical Accompaniment." *Proceedings of the 9th International Conference on Music Information Retrieval*, Philadelphia, pp. 57–62, 2008.

[7] T. Coughlan and P. Johnson: "Interaction in Creative Tasks: Ideation, Representation and Evaluation in Composition." *Proceedings of ACM Conference on Human Factors in Computing Systems*, Montréal, Canada, pp. 531-540, 2006.

[8] M. M. Farbood, H. Kaufman, and K. Jennings: "Composing with Hyperscore: an Intuitive Interface for Visualizing Musical Structure." *Proceedings of International Computer Music Conference*, 2007.

[9] D. Griffiths: Radiohead's OK Computer (Thirty Three and a Third series). Continuum, New York, 2004.

[10] M. Hickey: "More or Less Creative? A Comparison of the Composition Processes and Products of 'Highly-Creative' and 'Less-Creative' Children Composers." www.faculty-web.at.northwestern.edu/music/hickey/Exeter/Exeter.pdf.

[11] N. Hubbs: "The Imagination of Pop-Rock Criticism." *Expression in Pop-Rock Music: Critical and Analytical Essays* (2nd edition). Routledge, New York, pp. 215–237, 2008.

[12] A. Mardirossian and E. Chew: "Visualizing Music: Tonal Progressions and Distributions." *Proceedings of the 8th International Conference on Music Information Retrieval*, Austrian Computer Society, Vienna, pp. 189–194, 2007.

[13] M. T. Pearce and G. A. Wiggins: "Evaluating Cognitive Models of Musical Composition." *Proceedings of the 4th International Joint Workshop on Computational Creativity*, London, pp. 73–80, 2007.

[14] J. Tate: *Music and Art of Radiohead*. Ashgate, Aldershot UK, 2005.

[15] N. Vempala and S. A. Dasgupta: "Computational Model of the Music of Stevie Ray Vaughan." *Proceedings of the 6th Creativity and Cognition Conference*, Washington D.C., pp. 203–212, 2007.