Determining the impact of peer-feedback in lead sheet music composition

Daniel Martín, Benjamin Frantz, François Pachet
Sony CSL Paris
daniel.martin@csl.sony.fr

Abstract

To which extent peer feedback can affect the quality of a music composition? Is this impact related to the musical skills of the commentator, i.e. the person who does the feedback? To answer these questions we run an experiment in which each subject composes a short song using a lead sheet editor, then comments on someone else's song and finally the subject is asked to reconsider his initial song and try to improve it with the help of the feedback received. The results show that peer feedback can improve the quality of a song composition, especially when suggestions are well explained.

1 Introduction

Peer feedback is the practice in which one's work is revised by peers, specially in pedagogical contexts, where students in a class revise other students' exercises. This model is opposed to the traditional one in which all students' works are revised only by the teacher. Peer feedback is claimed to have many benefits [Rollinson, 2005] such as saving teachers' time as well as many other pedagogical positive effects [Sadler and Good, 2006]. With the increase of online learning communities and MOOCS [September, 2013], peer feedback is becoming more and more popular. Peer feedback is not only useful in pedagogical contexts, it can be also useful to apply in creative tasks. In music composition, collaborative composition has been addressed in several studies [Donin,]. There are online creative communities in which music is composed collaboratively by multiple users [Settles and Dow, 2013]. In this study we want to focus on the role of peer feedback in music composition, concretely in lead sheet music composition; i.e. composing a simple song with a melody and chord grid and writing it down in a lead sheet. The feedback provided by peers consists on suggestions of changes in certain parts of the lead sheet: certain notes or chords. This musical suggestions can be accompanied by a text explanation. Feedback can be checked by the composer, who can decide to use it or discard it. This is an anonymous one-way feedback with no dialogue, so it is different from the concept of collaborative composition where several composers work together hand by hand in a composition.

To which extent peer-feedback can affect the quality of a musical composition? To which extent the impact of the feedback is related to the musical skills of the commentator? In order to address this questions we have designed a music composition experiment.

2 Experiment

In this experiment subjects must write a short composition using an on-line lead sheet editor [Martín D., 2015]. Then they give feedback to another subjects' composition, and finally they are asked to reconsider their own original composition and try to improve it. Subjects are divided randomly in two groups: subjects in the control group (G1) are not receiving feedback, so they try to improve the song by themselves, whereas subjects from the experimental group (G2) will be able to use the feedback received to improve their own song. The existence of these two groups is ignored by the users so that the results are not biased.

As we are trying to assess the impact of feedback in the quality of a music composition, we need to evaluate the quality of all the compositions. We rely on social consensus to determine the quality of a song, i.e. the more listeners like it, the higher the quality is. Therefore, subjects will listen and evaluate other subjects' compositions. We describe each phase in the process of the experiment:

2.1 Questionnaire

Subjects start the experiment by answering some questions related to their experience in music, and also concretely in music composition. For example, they are asked how many years they have studied music theory, how many years they have been playing in a band, which style of music they like more, how often do they compose... etc.

2.2 Original composition

Subjects then write a short composition using the online lead sheet editor. A lead sheet is a particular type of music score widely used in jazz, bossa-nova and song-writing, consisting on a monophonic melody and a chord grid. All compositions have a fixed length of 8 bars; subjects are not able to add or delete bars. However, they can choose the tempo and the time signature. Subjects fill the 8 bars with a melody and chord labels (e.g. Dmaj7, Em7...etc.). Figure 1 shows a screenshot of the leadsheet editor.

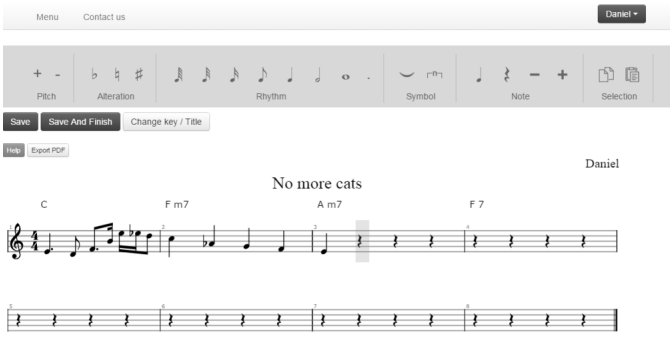

Figure 1: Screenshot of a composition being entered with the lead sheet editor.

Subjects can listen to their composition with the MIDI player. When they are done they click on "Save and Finish". Next, they answer a questionnaire about their confidence in the quality, complexity and satisfaction on their composition.

2.3 Suggesting

Once they have finished their composition they are asked to give feedback to another subject by suggesting improvements in another subjects' composition. Each suggestion will include a maximum of two bars. Subjects can make as many suggestions as they want as long as they do not overlap. So, each subject can make a maximum of 8 suggestions (one per bar). To do the suggestion, subjects must choose the bar(s) to modify, then they can change the notes and the chord symbols. Optionally, they can also leave a text comment explaining their changes. Figure 2 shows a composition in which a subject is entering suggestions with an explanation. When they are finished they answer a short questionnaire about their confidence on the suggestions they made as well as their opinion on the original song they modified.

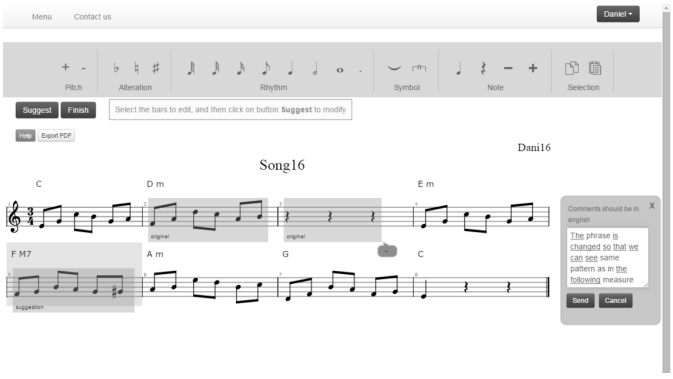

Figure 2: Screenshot showing the moment in which subject is entering an explanation of the suggestion.

2.4 Improvement: Final composition

Next, subjects are asked to reconsider their own composition and try to improve it. Subjects from G1 (group control) are told that they unfortunately did not receive suggestions and are encouraged to try to improve their own composition by themselves. On the other hand, subjects from G2 will see the suggestions they received from two other subjects. They can listen to all the suggestions. If they like a suggestions they can 'integrate' it, so that it is kept. In addition of integrating suggestions, they can modify freely their composition. Once they are finished, they answer a questionnaire about their confidence on their own improvement and on their opinion on the suggestions received.

2.5 Evaluation phase

Last step is to evaluate pairs of compositions from other subjects. Each pair of songs consist on the original song and the improved song. Subjects are supposed judge each song and to place it in a column with a legend from 0 ("I don't like it") to 100 ("I like it very much"). They have to evaluate at least 5 pairs in order to finish the experiment.

3 Results

In this section we describe in detail the results obtained from each phase of the experiment.

3.1 Population

56 subjects completed the experiment (68% men and 32% women). Mean age was 27.7 years, ranging from 19 to 61.

Musical experience was measured through a questionnaire with 7 items. The scale has a satisfactory sensibility with an observed range from 7 to 40 (out of 0 to 42) and we observed a mean of 27.9 (SD=8.8).The intern consistency is satisfactory (Cronbach's alpha=.82).

Composition experience was measured through a questionnaire with 5 items. The results show a low level of experience concerning composition in our sample with a mean 5.9 (SD=5.0) on a scale ranging from 0 to 30). The intern consistency is satisfactory (Cronbach's alpha=.83).

3.2 Composition effects

Each subject was randomly assigned to either the control group (G1) or the experimental group (G2). No significant differences were observed between the two groups in relation to age, gender, musical experience or composition experience; neither on the mean score evaluation (by other subjects) on the original song, nor on the composing duration which was tracked by the system (real composition duration) and also asked to the subject (self-evaluated composition).

Original Composition

The questionnaire that subjects were asked to complete after finishing the original composition included self-estimation questions about the quality, complexity and satisfaction for their composition on scales ranging from very bad/simple/unsatisfied (0) to very good/complex/satisfied (6). We also asked them to evaluate the time they spent to make their composition and if they used an instrument to help them to compose (and which instrument if they did).

Results show a mean quality of 2.9 (SD=1.6), a mean complexity of 1.9 (SD=1.7) and a mean satisfaction of 3.2 (SD=1.6). Only the complexity is significantly different to the center of the scales which is 3 ($t[43]=-3.39$; $p<.0001$). This means that the subjects tend to judge their work as rather simple (low complexity). We also observed positive and significant correlations between this three measures, ranging from r=.46 to r=.81.

During the suggestion step, we asked the subjects to also rate the quality and complexity of the songs they had to comment. Each composition from the experimental group (G2) was commented by two different subjects; so in the end we have the rating from the author and two other ratings from two different commentators. Interestingly, there was no correlation between the ratings from the original compositor and the ones from the commentators (—r—<.10), but the two commentators did agree together on the quality (r=.82) and on the complexity (r=.62).

Moreover, from the judgements done in the evaluation phase in which subjects evaluate pairs of songs from other subjects, the measurement of the quality of each song allows us to evaluate directly the real composition skills level of its author. Surprisingly, we observed that the quality of the original song is very slightly linked to the composition experience (r=.20, p=.14) or to the musical experience (r=.22, p=.10).

We also asked the subjects whether they used an instrument to help them in their composition. Results show a marginally significant effect in favour of the use of an instrument on the mean quality score ($t[54]=-1.69$, p=.10).

The mean duration of the composition time of the song as evaluated by the subjects is 31 minutes (SD=33 min) ranging from 1 minute to 240 minutes. This evaluation is largely underestimated by the subjects because the real duration calculated from the time spent on the composition software is significantly longer (m=64 min; $t[52]=3.54$, $p<.001$). The correlation between these two durations is not very high, but significant (r=.49, $p<.001$) indicating that the error of duration estimation is not exactly the same for everyone. Interestingly, we observed that the quality of the original songs (from the evaluation phase) is not linked with the time spent to compose, whether it is subjective (r=.05) or objective (r=.15).

Suggestions

In the questionnaire after making the suggestions, subjects were asked how much do they think the song they are revising will be improved due to their modifications (on a 7 points Likert scale ranging from 0 ”very little”, to 6 ”very much”).

The subjects from G2, the experimental group (N=25), received two suggestions to aid them for their final composition. Once they finished, we asked them if the suggestions received were interesting (on a 7 points Likert scale ranging from 0 ”very little”, to 6 ”very much”). Additionally, we recorded the number of suggestions they received and the number of texts comments received.

We ran a series of correlations between these measures and the improvement effect (the difference between the original song and the final song on the quality judgement score). None were significant, but we observed a significant correlation for the number of texts received (r=.56, p=.004). The correlation with the number of suggestions received is also positive (but non-significant r=.34, p=.09). We also have created two groups based on the quality level of the commentators which was used as an indicator of their composition skills resulting in a high skill commentator group (C+) and a low skill commentator group (C-). Interestingly, this composition skill level has no impact on the improvement effect ($t[22]=-1.61$, p=.12).

These results suggest that the raw quantity of information received is crucial to improve a composition whereas the music experience, composition experience or composition skills of the commentators are not.

Final composition

Overall, we can see that the control group, G1, does not improve significantly between the original song (m=54.6) and the final song (m=55.6) (improvement effect = .9, $t[31]=0.90$, p=.46). However, we do see a significant improvement for the experimental group, G2, between the original song (m=50.3) and the final song (m=53.1) (improvement effect = 2.8, $t[25]=2.86$, p=.04). See Figure 3.

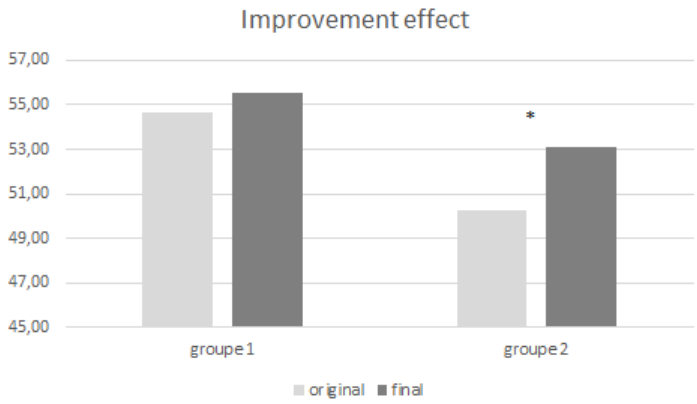

Figure 3: Difference between the original song and the final song on on the quality judgement score for the group without feedbacks (G1) and the group with feedbacks (G2).

We also examined the subjective evaluation of the participants concerning the improvement of their song. We constructed two composite scores. One from the self-evaluation scales of the original song (quality, complexity and satisfaction), one from the self-evaluation scales of the final song (quality, complexity and satisfaction) The intern consistency of those composite scores are satisfactory (the three Cronbach' alphas are above .82). We then conducted a mixed between (control and experimental groups) x within subjects (original and final song) analysis of variance. Overall, we observed a subjective improvement ($F[1,54]=12.84$, $p<.001$) meaning that both groups judged their final composition better than the original one. However, we also observed a significant interaction between groups and songs, ($F[1, 54] = 4.64$, p=.03), meaning that participants who received suggestions judged the improvement between the original and final song higher than the control group.

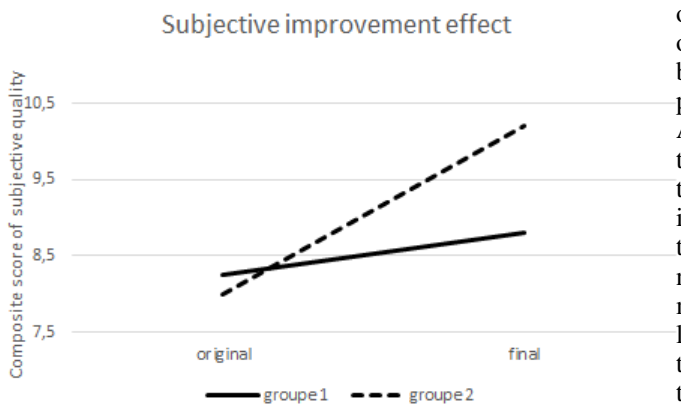

Figure 4: Self-esteemed quality of the original and final songs for the group without feedbacks (G1) and the group with feedbacks (G2).

Lead sheet editor

The software used was developed specifically for the experiment and we asked subject whether it was frustrating (0) or helpful (6) to compose with it. Results show a mean of 3.13 after the first composition and 3.40 after the final composition (the difference is not significant) which means that even if the online editor was not specially helpful, it did not hinder the composition process.

4 Conclusion

The aim of this experiment was firstly to examine the impact of peer feedback in music composition and secondly to determine which characteristics of the commentators are the most important. Before any improvement or suggestions, subjects had to write their first song. Interestingly, results show that subjects' previous experience in composition did not predict the quality of their song. The same pattern was also found for the subjects' previous experience as musician. These two results suggest that the quality of a song (based on social consensus) does not really tap in musicality but in something else, presumably creativity.

As expected, results show that anonymous peer feedbacks have a beneficial effect on music composition: only the participants who received feedbacks successfully improved their composition. This basic finding suggests that improvements in music may be achieved even without real collaboration with dialogues and active interactions, but by simple suggestions on a single occasion. These results stand whether the quality of a song is assessed by a social consensus or self-esteemed by its author. Hence, participants who received feedbacks not only feel that they have done a better song after the improvement step, but they actually did.

We also wanted to assess what makes a suggestion a good one. Results show that when a composer acts as a commentator by providing feedback to someone else's composition, neither the commentator's composition experience nor the commentator's composition skills (evaluated by the quality of their first song) have a positive impact on the improvement on the song. This finding is consistent with our results on the original song, and suggest that the quality of a song (again, based on social consensus) does not really tap in musical experience (as a musician or a composer) but in something else. As before, creativity might play a role, but also the clarity of the suggestions, because commentators had only one opportunity to make suggestions and explain them. This last point is confirmed by our results which show that the most important aspect of the suggestions is the quantity of text, i.e. the number of comments: The larger the number of text comments, the larger the improvement effect. This finding is in line with [Van den Berg *et al.*, 2006] who show that explanations of the content of the feedbacks (i.e. text comments on the suggestions), is crucial.

Acknowledgments

This work is supported by the Praise project (EU FP7 number 388770), a collaborative project funded by the European Commission under programme FP7-ICT-2011-8.

References

[Donin,] Nicolas Donin. Domesticating gesture: the collaborative creative process of florence baschet's streicherkreis for 'augmented' string quartet (2006-2008). *Eric Clarke & Mark Doffman (eds.), Creativity, Improvisation and Collaboration: Perspectives on the Performance of Contemporary Music, New York: Oxford University Press*.

[Martín D., 2015] Pachet F. Martín D., Neullas T. Leadsheetjs: A javascript library for online lead sheet editing. In *First International Conference on Technologies for Music Notation and Representation (TENOR)*, Paris, France, 2015.

[Rollinson, 2005] Paul Rollinson. Using peer feedback in the esl writing class. *ELT journal*, 59(1):23–30, 2005.

[Sadler and Good, 2006] Philip M Sadler and Eddie Good. The impact of self-and peer-grading on student learning. *Educational assessment*, 11(1):1–31, 2006.

[September, 2013] On September. Behind the scenes with moocs: Berklee college of musics experience developing, running, and evaluating. *CONTINUING HIGHER EDUCATION REVIEW*, 77:137, 2013.

[Settles and Dow, 2013] Burr Settles and Steven Dow. Let's get together: the formation and success of online creative collaborations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2009–2018. ACM, 2013.

[Van den Berg *et al.*, 2006] Ineke Van den Berg, Wilfried Admiraal, and Albert Pilot. Designing student peer assessment in higher education: Analysis of written and oral peer feedback. *Teaching in Higher Education*, 11(2):135–147, 2006.