Chapter 4

Evaluation

This chapter evaluates Excello against the success criteria and by showcasing examples. Converting of MIDI corpora to Excello using the converter demonstrates the notation's expressivity. Next, the summative evaluation user study involving 19 participants is explained and its data used to assess the features implemented in the participatory design process and to reason about Excello using the CDN framework [11]. Finally, ethics and data handling procedures are covered.

4.1 Excello Successes

A musical notation and program integrated into Excel for playback have been implemented. As in the success criteria, users can play multiple notes and chords of different durations. These can be combined into looped sequences and have defined tempo. Additional features were added as extensions; multiple successive notes in a cell, turtles automatically moving forward, and nested instructions with repeats all facilitate more efficient notation. Custom Excel functions, a chord adding tool and faster "turtle toggling" allows users to work more efficiently. Figure 4.1 shows Excello in use with a participant's arrangement.

The first section of Reich's Piano Phase is two identical piano melodies, one played slightly faster than the other [8]. The parts move out of phase, periodically aligning at different offsets. This is included as an example for many end-user programming tools, perhaps as it cannot be concisely notated by western staff notation. Manhattan implements it using three rows of 24 columns [21]. Sonic Pi requires one line for the notes and eight for playback. Excello only requires two cells to define two turtles with different speeds in addition to the notes. All three implementations are shown in Figure 4.2.

4.2 MIDI Corpus Conversion

Whilst concisely notating some music that western notation cannot, Excello can also exactly express pieces defined in MIDI. Tempo being redefined within a track is not be accounted for. If instead the time between messages is adjusted, the uncompressed conversion accounts

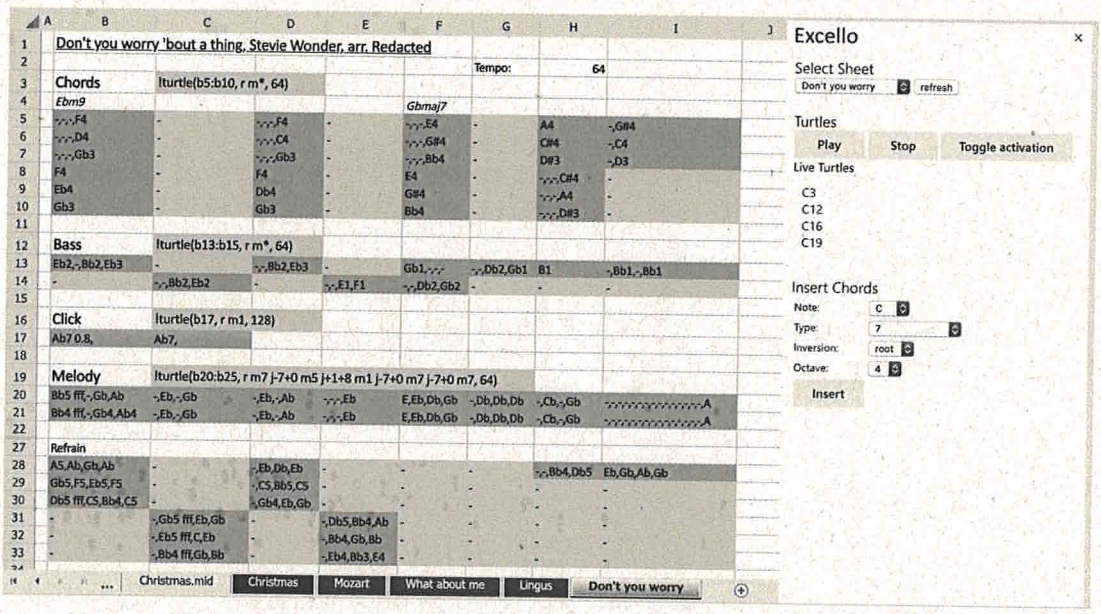
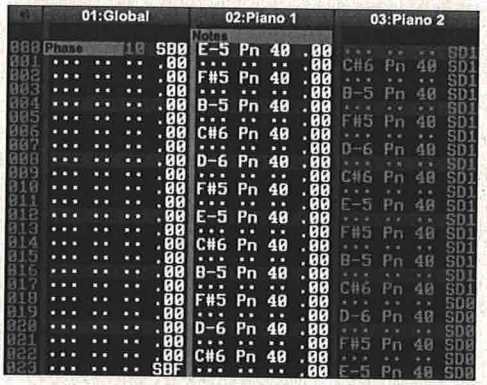


Figure 4.1: An arrangement with separated and labelled parts per instrument. Turtles refer to a global tempo at the top of the spreadsheet.



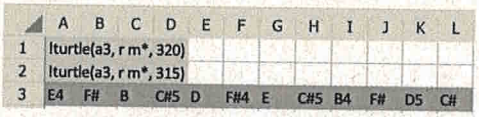
(a) Column 01 keeps track of the phase, 02 defines the notes and 03 is the phased notes - defined with formulae taking the phase and defined notes

```
notes = (ring :E4, :Fs4, :B4, :Cs5, :D5,
              :Fs4, :E4, :Cs5, :B4, :Fs4, :D5, :Cs5)

live_loop :slow do
  play notes.tick, release: 0.1
  sleep 0.2
end

live_loop :faster do
  play notes.tick, release: 0.1
  sleep 0.195
end
```

(b) The defined notes are played by two concurrent loops with different gaps between each note.



(c) Two turtles play the same notes at different speeds.

Figure 4.2: Reich's Piano Phase in (a) Manhattan, (b) Sonic Pi, (c) Excello

for this but the compressing algorithms will produce erroneous results as the difference between notes deviates too far from non-integer multiples of the minimum. Control messages like piano pedalling are not supported. Provided the difference between any two notes is a multiple of the minimum difference, the compression method that divides times by this accurately reproduces the MIDI, resulting in spreadsheets orders of magnitude smaller. This method would not accurately convert quavers against triplets (three notes played in the same time as two) provided these notes were not integer multiples of a smaller note. Given that the lengths of MIDI notes can differ from the note's time in standard notation, an assumption on the ratio of note lengths was required for compression. The modal compressive conversion is lossy if the minimum note distance is not the modal

something odd

oh I think I know what you mean now, sorry, it's a bit confusing ÷ ×

ignore me! sounds like a made-up word!

Why do you define this in a bracket where you've used Δnotes previously?

distance. This is useful in pieces with infrequently occuring ornaments or notes that dramatically decrease the minimum distance. Therefore this loss may be useful for more efficient representations.

I have converted three MIDI corpora. A collection of 497 Bach chorales[1] made by Margaret Greentree, 280 piano pieces[2] held by Bernd Krueger under a creative commons license,[3] and 194 Bach pieces from "A Johann Sebastian Bach Midi Page".[4] This is not all the files from this site as some were not readable using the Python library Mido. All 971 MIDI files were converted using all three methods. The Excello language is sufficiently expressive to represent MIDI files and can do so concisely, provided the condition of minimum note onset differences is maintained.

4.3 Summative Evaluation Sessions

19 of the 21 participants continued using Excello after formative evaluation session and answered a summative evaluation questionnaire. First, the features added after the initial sessions were recapped. To ensure users sufficiently understood the interface before giving feedback, a short transcription task also requiring some authoring was given.

The questionnaire first evaluated the participatory design process by comparing the interface before and after each feature. Seven-point Likert scale questions assessed if the issues had been solved and if overall the system was preferable. The remaining questions were based on Blackwell and Greens' questionnaire [5] - a tool for evaluating information devices' usability using the CDN framework. For example, the dimension *Role Expresiveness* - how much an element suggests its purpose - is assessed by: "*Are there some parts that are particularly difficult to interpret?*". The questions used are shown in Table 4.4. CDN can be used to analyse musical notation [9] and software systems [12], so is suitable to discuss Excello's notation and interface. Dimensions' significance for different activities varies [11], so users identified the percentage of time they spent carrying out these activities (searching for information, translating, incrementation, modification and exploratory design). Likert scale questions focusing on closeness of mapping, consistency, secondary notation, viscosity and visibility were used as planned in the proposal. It was suspected that reasoning about cognitive dimensions would be more challenging for participants, so to reduce the expected variance, only a five-point scale was used. The two have been shown to produce similar results [7]. CDN results were also collected for the user's preferred music composition interface. 12 users chose Sibelius, which was used for comparison.

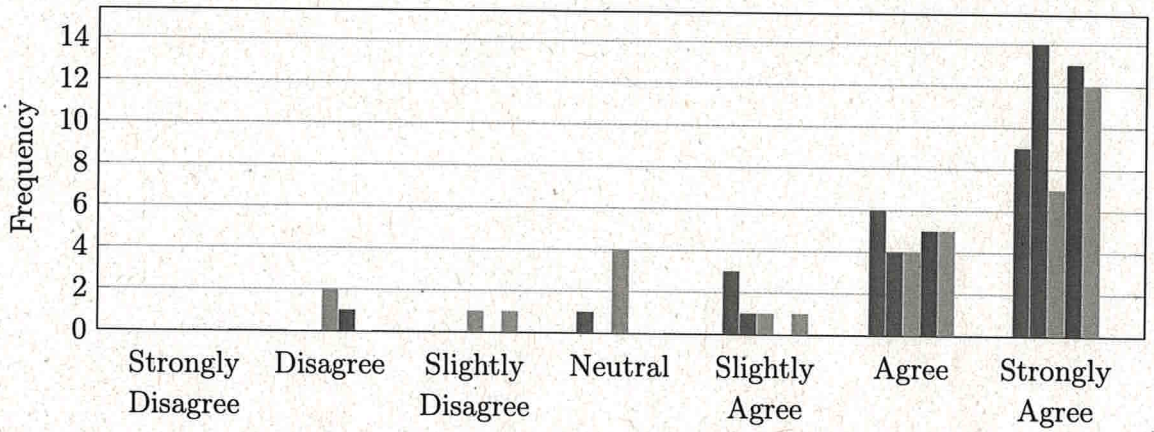
[1] Accessed from https://github.com/jamesrobertlloyd/infinite-bach/tree/master/data/chorales/midi.
[2] http://piano-midi.de/midis.htm.
[3] https://creativecommons.org/licenses/by-sa/3.0/de/deed.en.
[4] http://www.bachcentral.com/midiindex.html.

4.4 Success of Participatory Design

For each feature added, Excello with (system 2) and without this feature were compared. The following charts show the Likert scale responses for each question. I considered the mode of the Likert scale [3]. Chi-squared goodness-of-fit tests confirm the distributions are significantly different to uniform. All expected values must be greater than 1 and 80% greater than or equal to five [23]. As the expected frequency for one result is $19/7 \approx 2.7$, I combine Strongly Disagree with Disagree, Strongly Agree with Agree and the remaining three options into a third group. The p-value from a chi-squared test with these categories is given.

4.4.1 Dynamics in the Cell

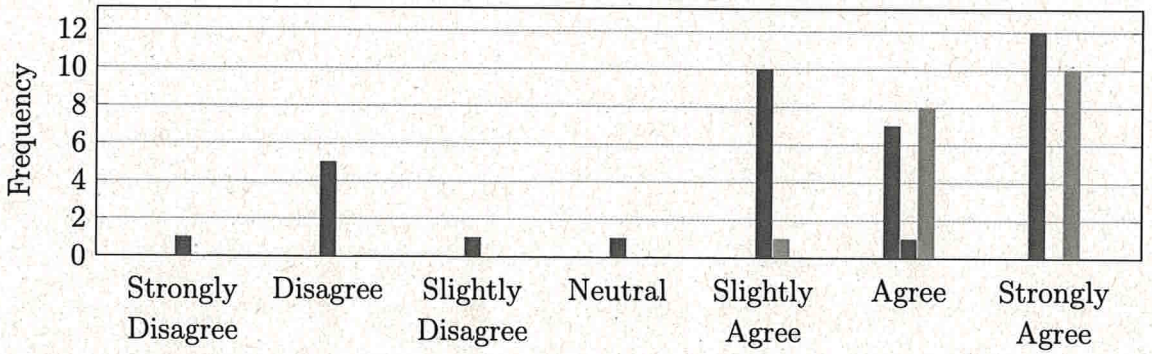
Statement	Mode	p-value
■ It is easier to figure out the turtle's path.	Strongly Agree	0.0000
■ It is easier to figure out what dynamics different notes are played at.	Strongly Agree	0.0146
■ It is easier to tell the order in which dynamics are applied.	Strongly Agree	0.0000
■ It is easier to write dynamics in the correct place.	Strongly Agree	0.0000
■ Overall system 2 is preferable.	Strongly Agree	0.0000



Volume instructions are defined in cells with notes rather than in turtles with movement instructions. There is strong evidence suggesting this change improved some of the issues identified during participatory design, resulting in an improved system.

4.4.2 Inferred Octave

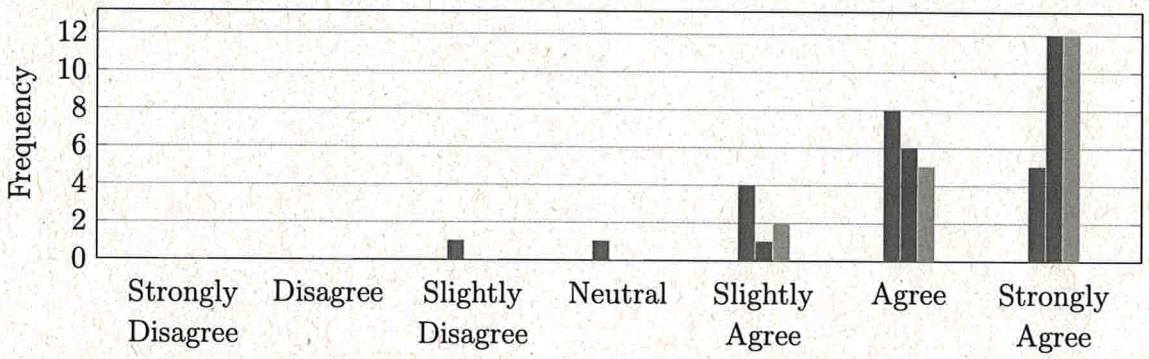
Statement	Mode	p-value
■ Less effort is required to write a part.	Strongly Agree	0.0000
■ It is harder to figure out what octave a note will be played in.	Slightly Agree	0.0639
■ Overall, system 2 is preferable.	Strongly Agree	0.0000



The inferred octave notation makes octaves harder to infer. However, the response distribution is not significantly different to uniform at the 5% level. Overall, this addition was significantly preferable.

4.4.3 Nested Instructions

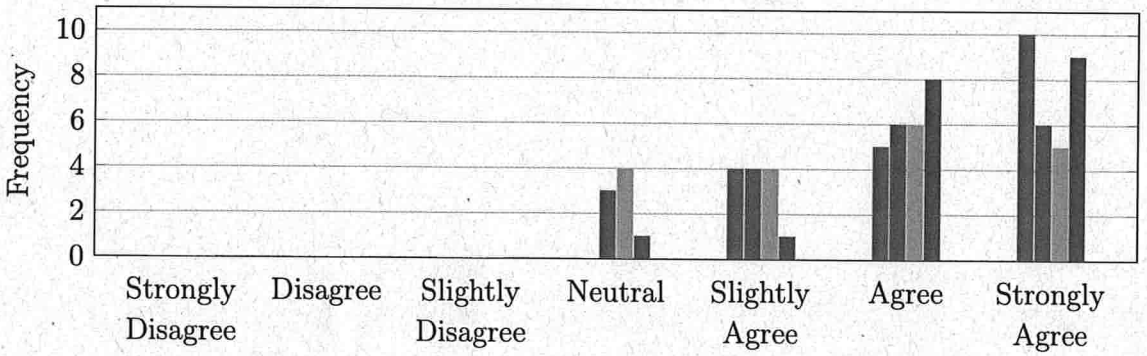
Statement	Mode	p-value
■ It is easier to parse the turtle instruction and tell what it will do.	Agree	0.0003
■ It is easier to repeat sections of notes.	Strongly Agree	0.0000
■ Overall, system 2 is preferable.	Strongly Agree	0.0000



All participants found the addition of nested instructions with repeats preferable, with the majority strongly agreeing.

4.4.4 Active Turtles List

Statement	Mode	p-value
■ It is easier to tell if a certain turtle has been registered.	Strongly Agree	0.0000
■ It is easier to see where the active turtles are.	(Strongly) Agree	0.0011
■ It is easier to toggle the activation of turtles.	Agree	0.0038
■ Overall, system 2 is preferable.	Strongly Agree	0.0000

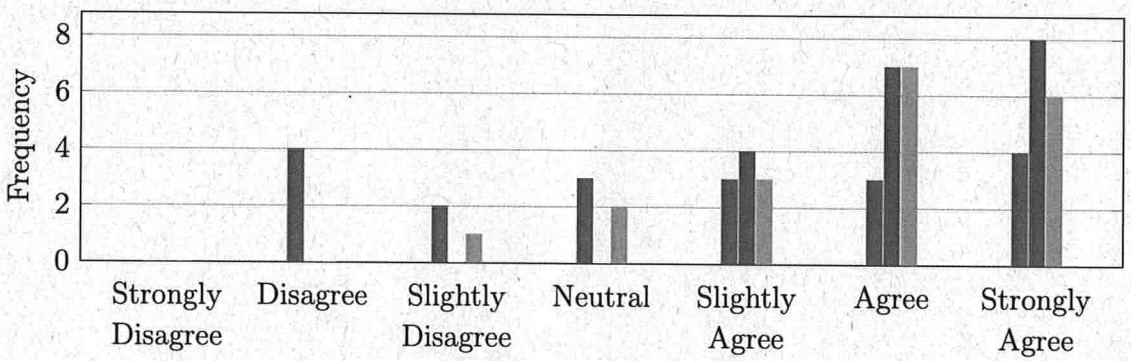


A list of active turtles had a neutral or positive effect for all users.

4.4.5 Continuous Volume

The volume of a note is more intuitive
I realise you might not be able to alter the questions now)

Statement	Mode	p-value
It is more intuitive how loud a note will be played.	Disagree Strongly Agree	0.6592
The volumes available are less limited.	Strongly Agree	0.0000
Overall, system 2 is preferable.	Agree	0.0003

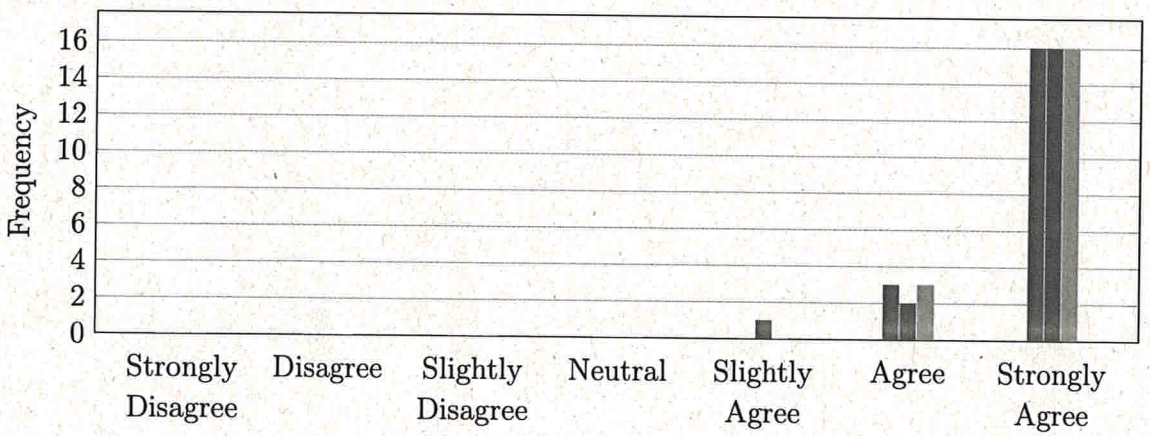


Isn't this a good thing? If so, why the "however"?

There is no significant result for whether the ability to define volume in the range [0,1] is more intuitive. All users agreed that the volumes were less confined. However, only one user did not find this change preferable. Given the previous conventional dynamic markings can still be used, this supports the addition being successful.

4.4.6 Automatic Stepping

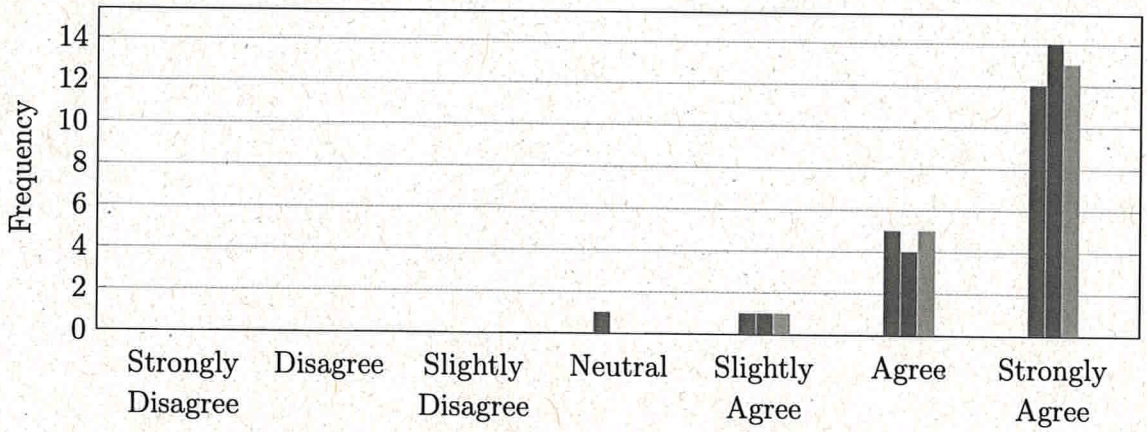
Statement	Mode	p-value
Less mental work is required to write the turtle instructions.	Strongly Agree	0.0000
Less work is required when more notes wish to be added.	Strongly Agree	0.0000
Overall, system 2 is preferable.	Strongly Agree	0.0000



This feature was particularly successful with 16 of the 19 users strongly agreeing the system was more preferable with automatic stepping available in the turtle instructions.

4.4.7 Absolute Tempo

Statement	Mode	p-value
It is easier to tell what the speed instruction corresponds to.	Strongly Agree	0.0000
Giving an exact tempo (e.g. when transcribing sheet music) is easier.	Strongly Agree	0.0000
Overall, system 2 is preferable.	Strongly Agree	0.0000



The initial design was changed so that turtle speed was defined in absolute cells per minute. All users found this to be an improvement.

Overall, the participatory design process was successful. Multiple features were added to Excello to solve problems identified through formative evaluation sessions and longer-term user feedback. There is evidence that all added features improved Excello.

4.5 Cognitive Dimensions of Notation

Figure 4.3 shows the time users identified carrying out the different cognitive activities [11] in Excello and in Sibelius. There are 19 Excello users and 12 for Sibelius. This shows translation is very important for both interfaces. There is more exploratory design for Excello but as users become more familiar with the system and more Excello notation exists, modification and incrementation may become more important. Little time is spent searching for information in the notation in either.

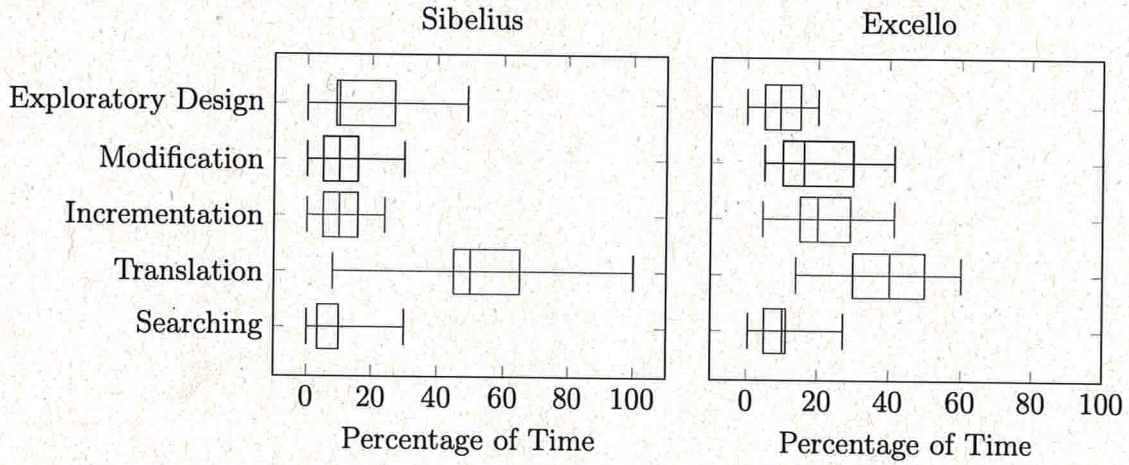


Figure 4.3: The percentage of time spent performing the different cognitive activities.

A series of statements from [5] were selected to assess the CDN of Excello. Users responded with a five-point (Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree) Likert scale. The significance of the results was verified with a chi-squared test. First, the data was combined into negative and non-negative categories. For each statement, the chi-squared test p-value and modal response are shown in Table 4.1. The distribution of responses is shown in Figure 4.4.

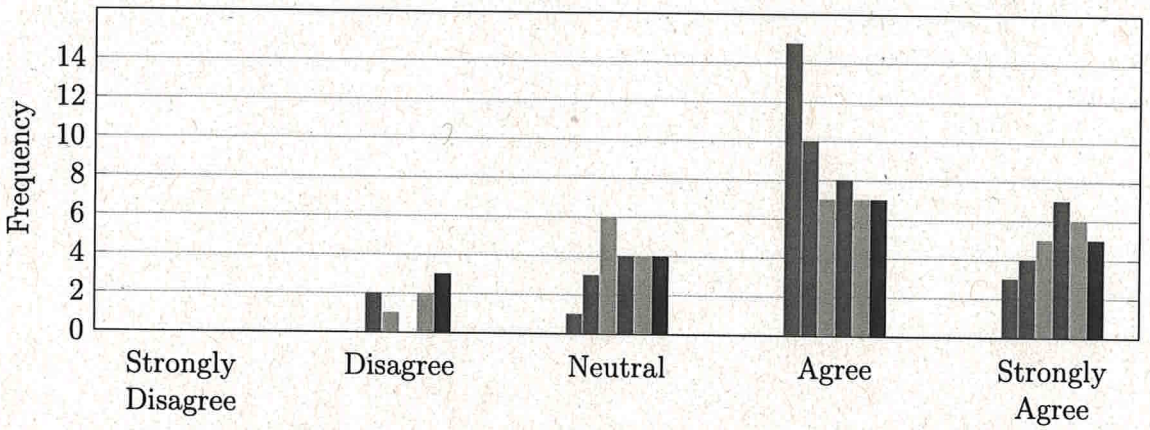


Figure 4.4: The responses to the questions in Table 4.1

Statement	CDN	Mode	p-value
(a) The notation used (In Excello: notes/ dynamics in cells and the definition of turtles) is related to the result you are describing (In Excello: Musical output)	Closeness of Mapping	Agree	0.0004
(b) Where there are different parts of the notation that mean similar things, the similarity is clear from the way they appear.	Consistency	Agree	0.0087
(c) You can add extra marks (or colours or format choices) to clarify, emphasise or repeat what is there already.	Secondary Notation	Agree	0.0020
(d) When you need to make changes to previous work, it is easy to make the change.	Viscosity	Agree	0.0004
(e) It is easy to see or find the various parts of the notation while it is being created or changed.	Visibility/ Juxtaposition	Agree	0.0087
(f) If you need to compare or combine different parts, you can see them at the same time.	Visibility/ Juxtaposition	Agree	0.0312

Table 4.1: Questions and results for testing the CDN of Excello

As these questions were also answered for the user's preferred interface, a comparison to Sibelius is made. As the data does not meet the assumptions of the t-test [3], I performed a Wilcoxon matched pairs signed-ranked test on the 12 pairs by encoding the five responses as -2,-1,0,1,2. For all six questions, there is no indication that the answers for the two interfaces come from populations with different means.

4.5.1 Closeness of Mapping

A test value of 5 for 5 changed pairs provides no significant evidence that the population means for Excello and Sibelius were different, suggesting Excello's notation with spreadsheets has not compromised the closeness of mapping of traditional notation. This is helped by using existing SPN for defining notes, the turtle instructions mapping to movement through the grid, and by the speed argument being an absolute, not relative, parameter. Being less familiar with staff notation, user 4 had found Sibelius's notation unintuitive.

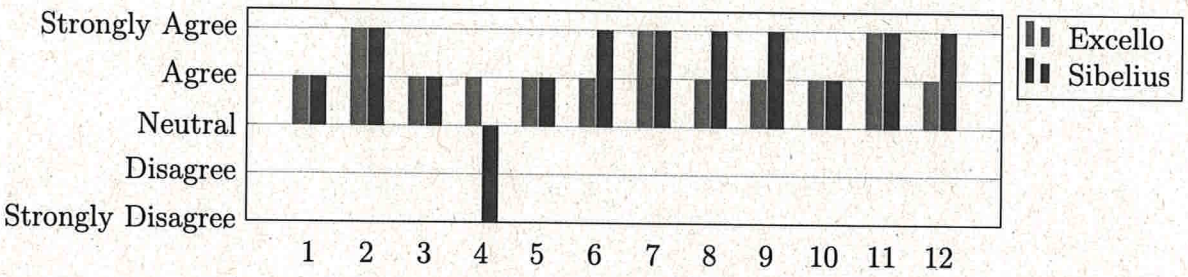


Figure 4.5: User responses for *closeness of mapping* for Excello and Sibelius from (a)

produces? creates? I don't remember reading causing notes up until now

4.5.2 Consistency

Each cell and turtle only causes one note at a time. Consistency is maintained by building pieces from these elements. Excello keeps consistency with Excel by sharing notations (e.g. A1:A5 for ranges) and using the existing formula editor. Using a number after instructions to repeat movements holds for both individual instructions and sequences. A test value of 12 for 7 changed pairs is not a significant result for the Wilcoxon test.

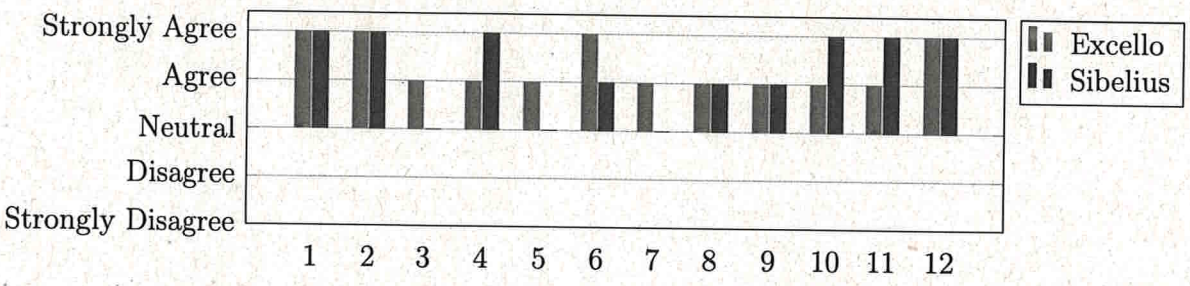


Figure 4.6: User responses for *consistency* for Excello and Sibelius from (b)

4.5.3 Secondary Notation

Given the time spent translating, secondary notation is particularly important [9]. As Excello abstracts time from the grid axes, the parts distribution is up to the user and cells can be used for arbitrary marks. Therefore, existing Excel features for formatting and grouping cells remain available. This is utilised by the automatic highlighting of notes and turtles. A test value of 15.5 for 8 changed pairs suggests no significant difference in population means. This suggests that the spreadsheet paradigm can provide equal secondary notation abilities to Sibelius, software already equipped with numerous ways to customise a score.

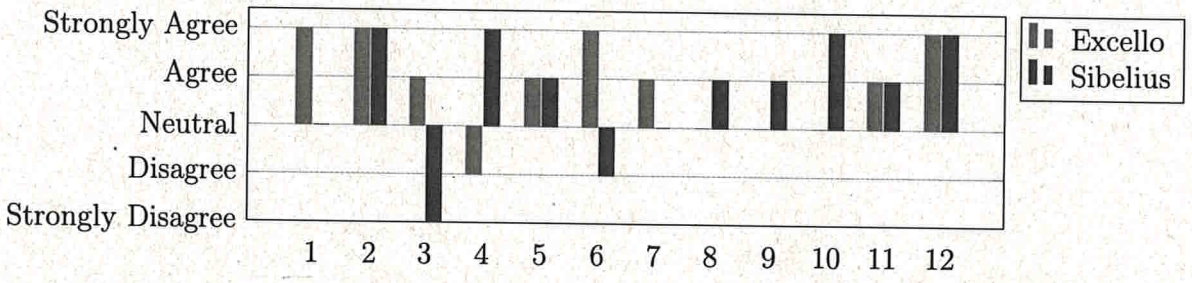


Figure 4.7: User responses for *secondary notation* for Excello and Sibelius from (c)

4.5.4 Viscosity

Allowing... to be ... and letting turtles...

Letting dynamics and octave marking be omitted and turtles stepping forward automatically provides low resistance to making additions and changes to the music. The toggle activation button dramatically reduces the actions to turn turtles on and off. Furthermore, Excel provides easy editing and movement of cells. A test value of 10.5 for 9

number of ?

changed pairs is not a significant result. This suggests the interfaces have comparable viscosity.

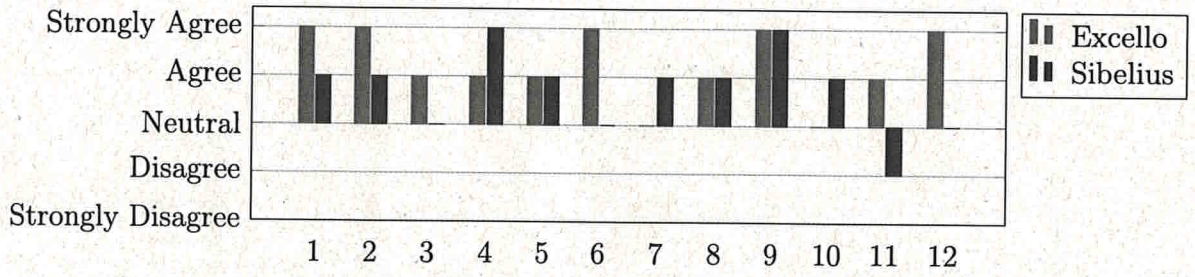


Figure 4.8: User responses for *viscosity* for Excello and Sibelius from (d)

4.5.5 Visibility / Juxtaposition

For both questions, there was no significant difference in population mean with test values of 6 and 7 for 6 and 7 changed pairs. This suggests that the spreadsheet interface can provide a similar ability to view components to Sibelius.

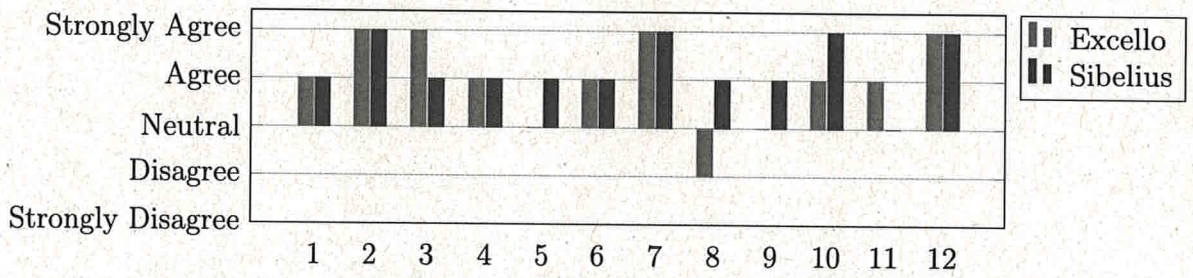


Figure 4.9: User responses for *visibility/juxtaposition* for Excello and Sibelius from (e)

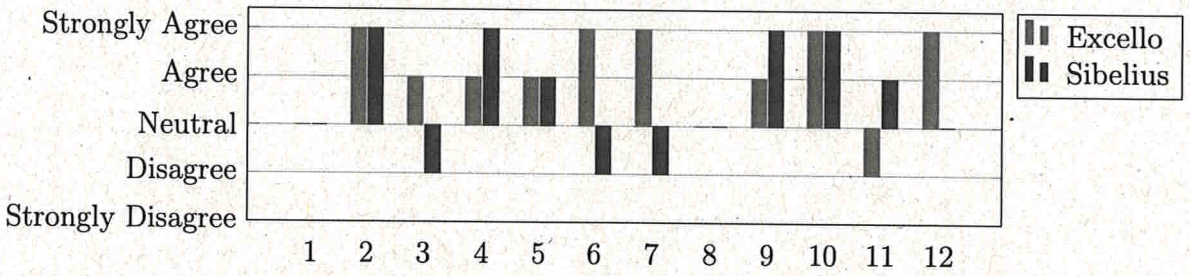


Figure 4.10: User responses for *visibility/juxtaposition* for Excello and Sibelius from (f)

Sibelius uses established music notation as part of professional software. However, there was no significant evidence to suggest Sibelius outperformed Excello across the CDN. This suggests that despite being a general purpose spreadsheet environment, Excello is a successful interface for writing music.

4.5.6 Other Dimensions

If users are unfamiliar with the turtle paradigm, this may reduce the *role-expressiveness.* But turtles and notes are the only musical spreadsheet components and are identified by highlighting. Whilst notes and turtles can be added in any order, adding parts may require many line insertions, increasing the *premature commitment.* The dual-formalism of turtles and notes could create high *diffuseness* but the layout flexibility allows this to be minimized as in Figure 4.1. This also shows how representations can have strong *synopsie*, as the notes or turtles don't need to be examined to understand what is happening. This may come at the expense of *hidden dependencies* if it is not immediately clear which notes are triggered by which turtles. Volume is also dependent on the notes turtles play before it. But as a single cell could be played at multiple volumes, this is a tradeoff of this design decision.

As well as the "`m*`" notation decreasing *viscosity*, it also improves the *progressive evaluation* as turtles can be played before a whole part has been transcribed. The ability to define a turtle and fill in the notes later also improves the *provisionality.* "`m*`" also reduces *hard mental operations* and the chord input tool removes manual calculation of the notes of chords.

Spreadsheets are "an abstraction-hating system" [11], therefore little *abstraction* is provided by Excel. Grouping turtles in one definition and nested bracketing of movement instructions improve this. These features also provide good *legibility.*

4.6 Ethics and Data Handling

After ethics approval, the pilot formative evaluation session was designed. After the pilot (also performed for summative evaluation), the session was revised before continuing with the remaining sessions. Participants filled in a consent form explaining the project and the session format. Participants had the choice to remain anonymous and not appear in acknowledgements. All participants' data included a unique ID to use to request removal or anonymising of their data. All participant data was only seen by myself. To prevent the jotting down of notes causing delays, formative evaluation sessions were audio recorded, typed up after the session, and then the audio was deleted. All participant data was also backed up on GitHub with the rest of the project but in an encrypted folder. Physical backups were also encrypted.

Chapter 5

Conclusion

The project set out to explore the hypothesis that spreadsheets would provide a productive medium for musical expression. Excello is a notation and corresponding program for musical playback integrated within Microsoft's Excel. By abstracting time away from the axes of the grid, the existing functionality of Excel remains highly useful. Having satisfied the initial success criteria for the program, development continued, carrying out participatory design with 21 users. As a result of this, many additional features beyond the initial scope of the project were implemented, all of which have been shown to significantly improve the interface. With respect to CDN, reasonable closeness of mapping, high consistency, high secondary notation, low viscosity and high visibility were all achieved as desired. Quantitatively, Excello is able to express a substantial subset of all MIDI music, and a converter was built to translate existing corpora of MIDI files to CSV files for Excello. The converter included two additional compression mechanisms, which still represent all musical information under certain, but common, conditions.

(commas ok but became v. choppy as the one I've added is necessary

During development, I submitted part of my code as an improvement to the open-source library Parenthesis. This was merged and has been published. The package has over 20,000 weekly npm downloads.

Excello freely provides a simple, but powerful program for musical composition to the hundreds of millions of users already familiar with the spreadsheet interface.

Word Count: 11745

What's the limit? Do you need help cutting out?