Perceptual preferences for groove patterns in jazz

Carl Leandersson
carllea@kth.se

Thorhildur Asgeirsdottir
tasg@kth.se

Pedro Campos
goncaloc@kth.se

Yan Yu
yan8@kth.se

Supervisor
Anders Friberg

Abstract
The experience of listening to jazz is impacted by inherent characteristics in the music. Some of these distinctive elements have to do with rhythm, harmonics, swing ratio, or correspondence between different instruments and players within an ensemble. These key features of the music come together to form a holistic experience, where each component has an effect on the resulting perception of the musical event. The general aim of this research is to investigate listeners' preferences for various patterns in jazz music samples, specifically the delay between soloist and drummer and how groovy that delay is perceived by the listener. This report outlines experimental results of testing for these differences amongst experienced listeners. Two samples were prepared and formatted in two different tempi with five variations of delay, which accounts for sufficient complexity of listening experience. The results show that, in line with prior research, participant's prefer microtiming that is close to the original performance.

1. Introduction

A key feature of jazz is variance in synchrony between the soloist and drummer. On the downbeat, the soloist is delayed relative to the drums, but on offbeats (weak eighth notes between the beats) the soloist and drummer are generally in sync. Deviations from precise metronomic timing between the drums and soloist have been established in previous research as a central piece in groove (Janata, Petr & Tomic, Stefan & Haberman, Jason, 2011). This experiment sets out to assess which amount of delay for soloists is perceived as the most groovy by participants. The definition of grooviness is provided by previous research and is defined as follows: "the sensation of movement or wanting to move while listening to certain types of music" (Davies, Madison, Silva, & Gouyon, 2013).

Samples were prepared from two jazz solos from Charlie Parker's *Donna Lee*. The MIDI files were altered to reflect two tempi, 160 and 180 bpm, and five implementations of varying delays.

The experiment was conducted with 20 participants, all of which had some musical experience, who listened to each sample and were subsequently asked to indicate their preference for each sample.

Given the range of possibilities of groove patterns in jazz, and similarly the diversity of listener's preference, this study sheds light on the correspondence of grooviness and magnitude of delay.

2. Background

2.1. Groove

The term groove has been defined in previous studies with varying nuance as the sensation of movement or wanting to move while listening to certain types of music (Davies, Madison, Silva, & Gouyon, 2013). In other domains such as music theory and musicology, a central factor in groove is the deviation from exact metronomic timing between drummer and soloist (Janata, Petr & Tomic, Stefan & Haberman, Jason, 2011). The effect of these timing deviations and how they relate to the perception and preference of the listener is an exciting research area that has been the basis for several studies. Davies, Madison, Silva, and Gouyon conducted a study in 2013 that investigated the effect of microtiming on the perception of groove in typical rhythm patterns for Jazz, Funk and Samba. The authors found that systematic microtiming led to decreased groove, liking and naturalness ratings amongst the participants with one exception being a simple short-long shuffle Jazz pattern.

2.2. Swing

Swing is a common term and playing style in traditional jazz music. The typical rhythmic pattern of swing can usually be identified in an eight-note pattern where the odd notes are lengthened and the even notes are shortened. The duration ratio between these notes is known as the swing ratio (Friberg and Sundström, 1997). Previous work has identified a correlation between swing ratio and tempi of a song. When the tempi of a song are increased, the swing ratio decreases (Friberg and Sundström, 2002). Senn, Kilchenmann, Von Georgi and Bullerjahn conducted a study in 2016 that investigated the effect of expert performance microtiming on listeners' experience of groove in swing and funk music. The authors exaggerated the original timing deviations from recorded performances that consisted of drums and bass with several different magnitudes. They found that listeners preferred the magnitudes that were close or equal to the performances original microtiming pattern. They also found that the fully quantized performances with no microtiming were rated equally high on groove as the original performance.

2.3. Previous research

Microtiming deviations in jazz performances have been researched before. Datseris, Ziereis, Albrecht, Hagmayer, Priesemann, and Geisel

conducted a study in 2019 where the authors manipulated 12 recordings of jazz standards, performed by a soloist. The soloist played the piano while listening to quantized bass and drum tracks. The recordings were then manipulated where the timing deviations were quantized, exaggerated and inverted. The participants, consisting of jazz musicians of various degrees, rated the quantized performances slightly higher and the exaggerated performances were rated lower than the original with regards to swing. Inversion showed no impact. The study concludes that naturally fluctuating microtiming deviations are not an essential factor for swing. Although the study assessed listeners preference in regards to swing, they did not assess grooviness.

3. Method

In this section, the methodology used for data collection is presented. Furthermore, the process of generating the music samples is described.

3.1. Participants

In total, 20 participants (all naive, 5 females, ages 21-61) took part in the experiment. Participants were recruited by research group members and were chosen based on prior musical experience either in general or in jazz specifically. Priority was given to those with more experience as lack of musical background could present a confounding factor in the results, as varying musical experience affects the listener's preference (Davies, Madison, Silva, & Gouyon, 2013). As part of the experimental testing, the participants were asked detailed questions about their specific experience and background with music, instruments, theory, and subjective listening experience.

Participants skewed toward younger age, with 95.2% of participants between the ages of 21-27.

Participants reflected a diverse range of nationalities: American, Austrian, Chinese, Dutch, Icelandic, Portuguese, Romanian, and Swedish. Whether nationality played an important role in the results is hard to determine at this time. It would be interesting to dig deeper into that question in further research, as well as whether other cultural factors might have an effect on preference of delay.

A total of 70% of participants had experience playing an instrument, with duration lasting from 6 months to 17 years. 45% of participants reported having experience playing in ensembles or organized musical groups. 60% of participants have studied music theory with the majority studying upwards of 4-16 years.

Participants' preference for music genres is wide ranging, with top genres being (in order) electronic music, pop music, jazz, rock, hip-hop, and funk.

When asked specifically about listening experience, 30% of respondents identified as extremely picky with music, i.e. not being tolerant of music they didn't like. The other 70% identified as having defined musical preferences, but not minding listening to a wide variety of music

even if it is music they don't particularly like.

Participants were probed on what they pay most attention to while listening to music, 60% answered they focus most on rhythm; 90% focus on melody; 60% focus on harmony; and 5% of participants reported not paying specific attention to any of these factors.

3.2. Music samples

For this study, two audio segments belonging to different jazz solos were used. These solos were selected based on simplicity, clarity and grooviness and were obtained by downloading the midi file for the track "Donna Lee" by Charlie Parker from the website notz.com, Jazz Midi.

The music editing software used to chop and shape the song was the paid version of Ableton: Music production with Live and Push.

Using this tool, all of the notes were fully quantized and synchronized to fit perfectly in time. This includes all the bass, piano and solo instrument notes. Then, 2 different tempi (160 and 180 bpm) were selected. The slower tempo was calculated using the original tempo of the song (180 bpm) and subtracting 20 bpm. This specific tempo was chosen since it still sounded well considering the musical context and also because there is research implying that 20 bpm is already enough difference to considerably affect the human perception of certain elements in jazz (Friberg, A. & Sundström, A., 1997).

The swing ratio is the duration ratio of the long-short pattern performed by the drummer. After doubling the amount of music segments from 2 to 4, the drums were added manually. Two different ride swing ratios were used: 2.75 for 160 bpm and 2.25 for 180 bpm. These swing ratio values were inspired by our supervisor's paper (Friberg, A. & Sundström, A., 1997) where they study the swing ratio value, across different tempi, contained in the strokes of expert jazz drummers.

Furthermore, the on-beat and off-beat notes of the solo melody were separated into two different tracks and the off-beat notes were synchronized with the drums to have the same swing ratio. The on-beat notes were delayed using steps of 18 and 16 ms for 160 bpm and 180 bpm, respectively, which correspond to the same amount of delay in terms of time compasses However, because 180 bpm produces a faster sample, the delay is also 2ms faster, which doesn't invalidate the fact that we're talking about the same amount of delay in terms of rhythm. This interval was readjusted and recalculated a couple of times so that the difference between samples was noticeable but also to keep a fine grain on the delay, for the purpose of this experiment.

In the last stage, 5 final samples for each music segment were generated by delaying five times the original music segment, producing a total of 20 samples combined.

	160bpm	180bpm

1 #	18ms	16ms
2 #	36ms	34ms
3 #	54ms	50ms
4 #	72ms	66ms
5 #	90ms	82ms

Table 1: Delay applied to each sample (ms) according to its tempo (bpm)

3.3. Questionnaire

Data was collected through a questionnaire format, where both demographic questions, music samples and corresponding preference scales were included.

The main purpose of our demographic questions was to get at participants' different backgrounds and musical experience. We asked about experience with musical instruments, ensembles, music theory and general musical preference such as favorite genre and pickiness. Generic questions about gender, age, and nationality were included as well to further probe differences in listening experience. The hope was that this extraneous information would give us a better sense of listeners' holistic experience of listening to the samples and hopefully provide insight into the different preferences exhibited in the results.

Concluding the demographic section, we presented the music samples organized by sample, tempo and delay. The participants were asked to listen to the samples and then rate their perception of its' grooviness, naturalness, and indicate their enjoyment of the sample.

3.3.1. Collected information

Information was securely collected and stored in a spreadsheet, later analyzed by group members.

3.3.2. Rating scales

The purpose of this study is to investigate what delay of the soloist sounds the most groovy to the listener. This information was not disclosed to the participants as this might affect their response. When assessing groove perception, it's important to select the right dimensions for measurement. Previous studies in music perception have used a Likert scale as a measurement for groove (Senn, Kilchenmann, von Georgi, & Bullerjahn, 2016). Therefore, the same methodology was used in this study, with participants rating samples on a scale from 1 (not at all) to 9 (very much). The scale 1-9 was chosen to provide a wide range of options with the possibility of providing a neutral response of 5. Three measurements were used; Groove, Naturalness and Enjoyment. Groove was selected as the main measurement for the study since our aim is to find the most groovy delay setting. Unlike Senn, Kilchenmann, von Georgi and Bullerjahn's study, we were not investigating different genres. Therefore, we had no risk of our participants being genre biased when asked about grooviness. Naturalness and enjoyment were selected as an additional subjective measure as it has

been used successfully in several studies for music perception (Senn, Kilchenmann, von Georgi, & Bullerjahn, 2016; Davies, Madison, Silva, & Gouyon, 2013).

3.4. Procedure

Participants were recruited based on their musical experience and were then sent a link to the participant form and asked to complete the experiment in a quiet setting. Participants provided demographic information before completing the experimental procedure of rating the music samples.

The music samples were presented in random order to prevent any noise in the data that might have resulted from listening to the samples in consecutive order, least delay to most. The order of randomness was the same across participants due to constraints of questionnaire functionality.

4. Results

The tables below show the mean scores and the 95% confidence intervals for each delay measurement, categorized by song and tempo. The confidence intervals were calculated using the formula $\bar{x} \pm Z(a/2) * \sigma/\sqrt{n}$ where $a = .95, n = 20$ and $\sigma = \sqrt{\Sigma (x - \bar{x})/n}$.

4.1. Grooviness

Song 1	160bpm	180bpm
1 #	**6.5** ± 0.79	6.35 ± 0.71
2 #	6.25 ± 0.66	**6.4** ± 0.52
3 #	5.85 ± 0.55	5.65 ± 0.67
4 #	5.55 ± 0.81	5.6 ± 0.52
5 #	5.3 ± 0.80	5.65 ± 0.83

Song 2	160bpm	180bpm
1 #	**5.6** ± 0.86	5.2 ± 0.84
2 #	5.2 ± 0.92	**5.55** ± 0.96
3 #	5.55 ± 0.85	5.1 ± 0.83
4 #	5.0 ± 0.8	**5.55** ± 0.84
5 #	4.35 ± 1.07	5.05 ± 0.79

4.2. Naturalness

Song 1	160bpm	180bpm
1 #	**6.2** ± 0.98	5.8 ± 0.82
2 #	6.1 ± 0.86	**5.9** ± 0.77
3 #	5.5 ± 0.87	4.75 ± 0.92
4 #	5.05 ± 1.21	5 ± 0.80
5 #	5.2 ± 0.91	5.45 ± 0.99

Song 2	160bpm	180bpm
1 #	**5.3** ± 1.11	4.65 ± 0.85
2 #	4.9 ± 1.12	5.05 ± 0.83
3 #	5.15 ± 0.95	4.5 ± 0.97
4 #	4.75 ± 1.08	**5.45** ± 0.85
5 #	4.25 ± 1.23	4.85 ± 0.76

4.3. Enjoyment

Song 1	160bpm	180bpm
1 #	**6.35** ± 0.73	6 ± 0.75

2 #	6.1 ± 0.69	**6.1** ± 0.63
3 #	6.15 ± 0.59	5.4 ± 0.71
4 #	5.65 ± 0.95	5.3 ± 0.63
5 #	5.4 ± 0.85	5.65 ± 0.91

Song 2	160bpm	180bpm
1 #	**5.7** ± 0.92	4.8 ± 0.94
2 #	5.15 ± 1.05	5.15 ± 0.86
3 #	5.55 ± 0.9	4.75 ± 0.88
4 #	5.05 ± 1.0	**5.5** ± 0.86
5 #	4.65 ± 1.11	5 ± 0.78

Within each group of samples, t-tests were conducted in order to test the significance of the difference in means. For grooviness ratings for Song 1 in 160 bpm, the difference between the mean scores of #1 and #5 is significant. For 180 bpm, the difference between the mean scores of #2 and #4 is significant. This implies that the participants had a substantial preference for the highest rated sample over the lowest rated in these two groups (in both cases shorter delays, #1 or #2, were prefered in terms of grooviness). However, for Song 2, none of the score differences were significant. The mean naturalness and enjoyment scores also show no statistically significant differences within each group of samples.

When the samples were filtered by their musical experiences (i.e., having played an instrument, studied music theory, or having jazz as a favorite genre), the end results were similar to what we've had above, with the less delayed versions (delay #1 and #2) being the most preferred. Nevertheless, most of the differences in grooviness scores are not statistically significant, with the exception of the people who had jazz as a favorite genre – their preference for delay #1 for song 1 were significant.

Then we compared the grooviness scores between the two songs. In the 160 bpm range, the participants gave all the Song 1 versions a mean rating of 5.89, and the Song 2 versions got a 5.14. The difference in means between these two groups is significant. In the 180 bpm range, Song 1 got a mean rating of 5.93, and Song 2 got 5.29. The difference in means between these groups is again significant.

5. Discussion

5.1. Statistical significance

Overall, the obtained results weren't substantial since most of the differences in scores weren't statistically significant. This might have been for several reasons.

Firstly, the number of participants could have been higher than 20 which could have provided statistically stronger results, however, due to constraints in time and practical issues this was not possible.

Providing a neutral rating of 5 might have adversely impacted the results. The mean was very close to 5 and a significant number of participants chose 5 for certain trials. It might have been better to use a linear scale of 1-10 instead of 1-9, that way the participants would be forced to choose between 5 and 6, which are polarizing, i.e. not have a neutral middle-option.

Furthermore, having a range of five different delays might have been too much, the difference was very fine grained between each sample and might have caused participants confusion. On the other hand, this might have been a good thing as it forced participants to be particularly tuned to each sample to pick out the difference and observe their preference. Had there been fewer samples, the difference might have been so obvious to the extent that the results would have been more homogenous.

Another aspect to consider is also the fact that the questionnaire turned out to be quite long (20+ minutes) which can have influenced the answers of participants. We noticed that the ratings for the second song were on average worse than the first song and we suspect it might have been because participants were already tired after completing the ratings for the first song. This can also have affected them to choose more neutral answers. However, we do not know this for sure since the order of the songs rated didn't change between participants.

Also, because of the nature of the samples, participants might have rated naturalness on their opinions of the MIDI instrumentation, instead of the rhythm of the piano solo. This could have led to the naturalness scores being too close to each other, and the differences being insignificant.

Lastly, it would have been ideal to carry out the experimental procedure in a lab setting where environmental factors could have been standardized for all participants. This might have provided a more natural experience as well, as the research group could have facilitated more of the experiment and provided in-person feedback and guidance.

5.2. Means

Nevertheless, the means show us that the preferred delay for both songs using 160 bpm was the first one of 16ms. In every example the highest attributed score for each parameter (grooviness, naturalness and enjoyment) corresponded to sample number 1.

Also, for song 1 with 180 bpm, the means show us that the prefered delay was the one of sample number 2 (34ms). When it comes to naturalness and enjoyment, the prefered delay for song 2 with 180 bpm was sample 4 (66ms) but when it comes to grooviness, both sample 2 (34ms) and 4 (66ms) were preferred amongst participants.

Despite a lot of the results not being statistically significant we can see that participants preferred longer delays (34ms and 66ms) for the faster tempo

and shorter (16ms) for the slower tempo. This trend is something that is supported by scientific literature (Ellis, 1991).

Finally, the difference between the preferred delays for 180 bpm between song 1 and 2 could have been caused by differences in the melodic structure of each song. The second song contained a phrase of delayed triplets that could have had an influence on the preferred delay to be longer in this certain tempo.

5.3. Conclusion

Although the results were not sizable, there was a trend of grooviness being preferably perceived in samples with shorter delays, while slightly longer delays are prefered as the bpm of the song increases. This shows that microtiming has an impact on listener's preference which is comparable to previous studies in the domain. Throughout the experimental procedure, various factors of the process were identified that could have been done better as noted in the discussion above. Further research would be advantageous to see if different, more fine-grained results would arise. This is an important inquiry to pursue as it is clear that there are various elements of musical timing that influence listening experience.

References

Datseris, George & Ziereis, Annika & Albrecht, Thorsten & Hagmayer, York & Priesemann, Viola & Geisel, Theo. (2019). Does it Swing? Microtiming Deviations and Swing Feeling in Jazz.

Davies, Matthew & Madison, Guy & Silva, Pedro & Gouyon, Fabien. (2013). The Effect of Microtiming Deviations on the Perception of Groove in Short Rhythms. Music Perception. 30. 498-511. 10.1525/mp.2013.30.5.497.

Ellis, Mark C., 1991. An Analysis of “Swing” Subdivision and Asynchronization in Three Jazz Saxophonists, Perceptual and Motor Skills, 73, 707-713.

Friberg, A. & Sundström, A., 1997. Preferred swing ratio in jazz as a function of tempo. TMH-QPSR, 38(4), pp. 019-027.

Friberg, A. and Sundström, A., 2002. Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern. Music perception, 19(3), pp.333-349.

Janata, Petr & Tomic, Stefan & Haberman, Jason. (2011). Sensorimotor Coupling in Music and the Psychology of the Groove. Journal of experimental psychology. General. 141. 54-75. 10.1037/a0024208.

Senn, O., Kilchenmann, L., von Georgi, R., & Bullerjahn, C. (2016). The effect of expert performance microtiming on listeners' experience of groove in swing or funk music. Frontiers in Psychology, 7, Article 1487. https://doi.org/10.3389/fpsyg.2016.01487

The Site For Jazz, African Music and Beyond, (2002). https://notz.com/jazz_midi.htm