# Music 251 Final Project Report

*Christopher Relyea*

**Title:** Perceived Humanness and Aesthetic Preference for Basic Gain and Time Modulations in Computer-Generated Drum Grooves

## Abstract

Microtiming and timbral micro-variations are central to the "naturalness" of human-performed music, and prior work suggests that perfectly quantized computer-generated rhythms are perceived as mechanical and undesirable for their lack of such variations. In contemporary music production, however, drum parts are frequently created from repeated one-shot samples (the same audio excerpt repeated without modulation, e.g. every snare sound in a groove is the same audio sample with no changes), and music producers commonly apply simple "humanization" techniques—small random timing offsets (jitter) and gain changes—to reduce the displeasing "mechanicalness" that occurs as a result of repeated sampling. It remains unclear how effective these basic manipulations are for one-shot drum-machine samples that lack the natural timbral variability of acoustic performance. This study evaluated the perceived effectiveness of timing jitter and gain variation applied to computer-generated drum grooves built from repeated kick, snare, and hi-hat samples. In an online listening study, participants rated each loop on 7-point scales of perceived mechanicalness and aesthetic preference. It was hypothesized that timing jitter would reduce perceived mechanicalness more than gain variation, that gain variation would improve aesthetic preference more than timing jitter. Hypotheses also included that moderate levels would outperform high levels, and that the most effective model would include simultaneous timing and gain variation. Results showed that all manipulations increased perceived humanness relative to the baseline, but gain variation produced the largest increase in humanness and the highest preference, whereas high timing jitter substantially reduced liking. Combined timing jitter and gain variation did not consistently outperform gain-only conditions for preference. These findings suggest that for repeated one-shot samples, modest gain variation may provide a more reliable route to "humanizing" grooves than adding substantial timing jitter, which can cross a threshold into perceived sloppiness and reduced musical enjoyment.

## Introduction

There is a large body of music information and psychoacoustical research concerned with the tiny yet perceptible variations in timing and timbral properties of music that emerge from the presence of a human performer. The rhythmic properties of a musical excerpt, elements which may be colloquially referred to as "groove" have been analyzed in detail for a number of years. One foundational finding in the perception of human-performed rhythmic sounds is that the perceived timing of a sound, as contrasted to its literal onset timestamp, is a factor of both the properties of the sound itself (envelope, timbre, spectral content) and the timing. These properties are inseparable; altering a sound affects where listeners perceive it to exist in time. By extension, both sound modulations (envelope, etc.) and timing variations affect the perceived rhythmic properties of music (Danielsen et al., 2024). Another study which analyzed real audio from human performances and measured drum hit onset times supported the assertion that the "best" musicians perform imperfectly. It was found that the timing jitter measured between regular human drum hits may follow a quasi-Gaussian distribution and may not be accurately modeled by simple random noise (Hennig et al., 2011).

Previous experimental results suggest that listeners are perceptually sensitive to these human imperfections. Several studies (e.g., Hennig et al., 2011) reproduced drum grooves with human-like jitter and compared them to those which featured the same sounds quantized to exact beats. Study participants, on average, preferred the human-modeled variation, suggesting that perfectly accurate rhythms generated by a computer produce a mechanical, undesirable effect. Jakubowski et al. (2022), further present a measured aesthetic preference for "isochrony vs. synchrony" which may be tied to the listener's culture and background. There is also evidence to suggest that human-modeled imperfection in drum grooves causes increased mental stimulation; one study found that subjects who listened to beats with microtiming variations produced stronger mental images of a character/scene than those who listened to exact rhythms (Ayyildiz et al., 2025). Listeners are also surprisingly able to implicitly discriminate numerical values of the mean and variance of added microtiming (Kaplan et al., 2023).

In the world of music production, it is often quicker, cheaper, and more customizable for producers to add drums to a song by synthesizing a beat from pre-recorded samples, rather than performing and recording on an acoustic drum set. Many digital audio workstations (DAWs) like Ableton, Logic, or Pro Tools feature the ability to easily create a new groove from a massive library of samples by sequencing rhythms, typically through the MIDI protocol. By default, these tools typically quantize a user's created groove to exact beats. The mechanical sound that results from this quantization may be an intentional choice on the part of the producer (i.e. the electronic music genre), but if a more human-performed sound is desired, it is common for producers to manually offset beats by small amounts to produce human jitter. To produce sound alterations per repeated sample, which can be thought of as simpler versions of timbral modulations (like those suggested in Danielsen et al., 2024), an artist might also slightly change the playback gain of consecutive samples. Producers resort to these techniques to reduce what is colloquially known as the "machine gun effect", or the unnatural, mechanical sound of a repeated drum sample (the same audio sample played back multiple times in a row) when no modulation is added between hits. Some DAWs even provide a "humanize" button, which applies these changes automatically to an entire groove at once. Hennig et al. (2011) note that these premade tools typically apply white noise variations in timing jitter.

The aforementioned studies which used measured jitter properties to create grooves for subjective aesthetic evaluation by study participants have all used either human performed grooves (slicing a human drum performance into samples and moving them in time) or physically-modeled drum software instruments (which automatically add timbral adjustments to repeated hits beyond simple gain changes by modeling the physics of drum heads and cymbals to create highly accurate sounds with automatic humanization). It is unclear if previously discovered aesthetic preferences for humanlike variations in generated drum grooves apply to grooves generated with a small number of repeated, non-physically modeled samples, i.e. all snare hits in a four-bar groove are the same audio sample played back at different times. Moreover, it is unknown if aesthetic effects of added jitter would exist for samples that were not originally performed by humans (i.e. not a recorded snare hit performed by a real human drummer, but a synthesized sound from an electronic drum machine). Producers often apply quick humanization techniques for timing and gain to these one-shot sampled grooves, but the effectiveness of such methods in producing natural sounding drum tracks has not yet been measured.

The present study aimed to recreate the music production drum groove creation environment by synthesizing grooves with a small number of repeated electronic drum samples and applying humanization techniques for timing and gain to measure their perceived effectiveness. Participants were asked to listen to a variety of generated drum grooves using hi-hat, snare, and kick drum samples from a drum machine. Generated grooves were modified with different amounts of time jitter and gain variation.

This study tested how microtiming jitter (timing offsets) and gain variation (hit-to-hit level differences) affect two perceptual outcomes for one-shot, sample-based electronic drum grooves: (1) perceived mechanicalness and (2) aesthetic preference.

H1 (Naturalness / mechanicalness):

Timing jitter may reduce perceived mechanicalness more than gain variation, since there is more evidence in the literature referenced by this study that supports the importance of timing over gain variation.

*Prediction*: Adding timing jitter will shift ratings toward "more human / less mechanical" more strongly than adding gain variation alone.

H2 (Aesthetic preference):

Gain variation may increase aesthetic preference more than timing jitter. This is because it may be difficult to properly recreate timing variations measured in previous studies through this experiment. This may also be because the population of participants were selected from a music technology degree program and have generally extensive musical experience, so they may be more sensitive to timing variations as errors (rather than natural variation) than other listeners.

*Prediction*: Listeners will "like" grooves with modest gain variation more than grooves with timing jitter alone, especially at low-to-medium levels.

H3 (Interaction / combined humanization):

Listeners' perception of naturalness may depend on timing and gain variations in an additive manner, even if they are independently varied.

*Prediction*: Combined conditions will outperform single-factor conditions, but effects may be non-linear (e.g., "too much" jitter/variation becomes messy or unpleasant).

H4 (Optimal level):

There will be an optimal "middle" range of variation (particularly for timing). When variation is increased beyond a certain threshold, it will create the effect of human error rather than human naturalness; too much timing variations will result in drum hits that are noticeably off-beat, and too much gain variation will sound like an inconsistent and unskilled human player who creates vastly different sounds on every hit. In other words, excessive variations may indicate incompetence of the "musician" or incoherence in music for listeners rather than musical expression or human-like nuances.

*Prediction*: Medium jitter/variation will be preferred over high jitter/variation, with high levels potentially increasing perceived sloppiness and decreasing aesthetic preference.

## Methods

10 participants were selected from the student body of Stanford University's center for Computer Research in Music and Acoustics (CCRMA). All participants were above the age of 18 and reported to have non-damaged hearing. Through a short questionnaire, the participants indicated their level of musical experience through the number of years they have studied/performed music and offered a brief description of their musical background. The mean reported number of years of music experience was 12.7.

Samples were sourced from publicly available 44.1 kHz WAV file recordings of a Roland TR-707 drum machine (kick, snare, and closed hi-hat one-shots). Identical samples were used for each repeated instrument hit. Grooves were constructed of 4-bar loops at 115 BPM.

There were 2 distinct 4-bar patterns used. Pattern A (Figure 1a) is a straight, even groove with sixteenth note closed hi-hats and a classic backbeat snare/kick pattern. Pattern B (Figure 1b) features the same hi-hat pattern but with a slightly more syncopated snare/kick structure.

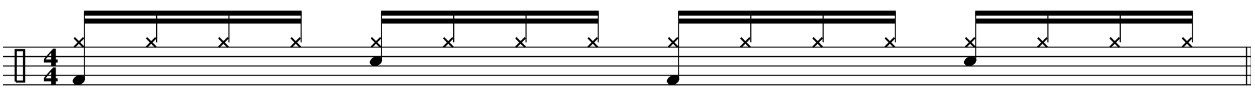

*Figure 1a. Pattern A*

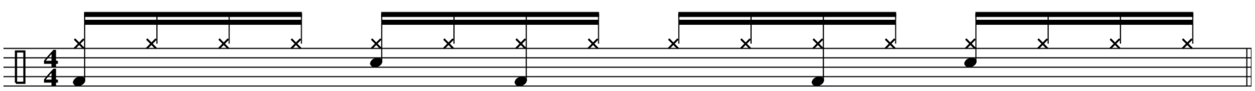

*Figure 1b. Pattern B*

The following humanization manipulations were applied to each groove:

1. Timing jitter: per-hit sample start (onset) offset jitter sampled from a zero-mean Gaussian distribution
   - Low: σ = 4ms
   - Medium: σ = 8ms
   - High: σ = 12ms
2. Gain variation: per-hit gain change sampled from a zero-mean Gaussian distribution
   - Low: σ = 1dB
   - Medium: σ = 2dB
   - High: σ = 3dB

To mirror typical DAW "humanize" tools, jitter and gain are applied independently per event and instrument.

*Conditions*

A 10-condition set was generated for each pattern.

*Table 1. Stimuli 10-condition set*

| **Condition** | **Timing jitter** | **Gain variation** |
|---|---|---|
| 1 | 0 | 0 |
| 2 | Low | 0 |
| 3 | Medium | 0 |
| 4 | High | 0 |

| | | |
|---|---|---|
| 5 | 0 | Low |
| 6 | 0 | Medium |
| 7 | 0 | High |
| 8 | Low | Low |
| 9 | Medium | Medium |
| 10 | High | High |

Total stimuli per participant:
2 patterns × 10 conditions = 20 loops

- Generation: Python (PyDub library for sound manipulation and WAV file construction)
- Timing resolution: Event offsets applied in milliseconds at waveform level.
- Clipping prevention: Applied normalization after rendering so gain variation doesn't clip.
- Export: Each loop exported as its own WAV file for Qualtrics embedding.

**Procedures**

Participants complete an online listening experiment (Qualtrics survey). They will listen to 20 short drum loops and rate each loop on two 7-point scales: mechanicalness and aesthetic preference.

Participants are instructed:

- Use headphones if possible.
- Sit in a quiet environment.
- Set volume to a comfortable level before starting using a test audio drum groove. They report:
    - Device type (laptop/phone/desktop)
    - Headphone model or speakers
    - Environment (quiet/noisy)
    - Any issues (dropouts, distractions)

For each stimulus:

1. Participant controls playback of audio sample (with replay allowed)
2. Participant answers two questions:
    - Mechanicalness**:** "How mechanical vs. human does this groove feel?"
      1 = Totally mechanical, 7 = Totally human/natural
    - Aesthetic preference: "How much do you like this groove?"
      1 = Strongly dislike, 7 = Strongly like

- Stimulus order randomized per participant

Duration: 20 loops × (≈10–15 seconds listening + rating time): expected total: 8–15 minutes

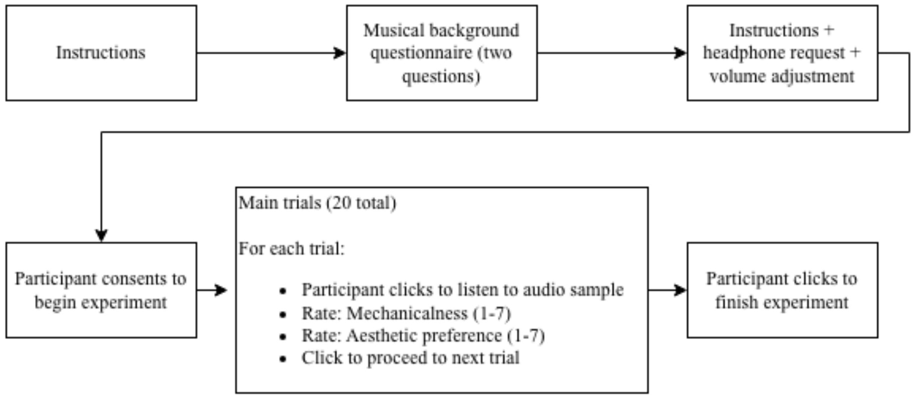


*Figure 2. Task timeline*

**Data Collection and Analysis**

For each participant/stimulus:

- Mechanicalness rating (1–7) (1=most mechanical, 7=most human)
- Aesthetic rating (1–7) (1=strongly dislike, 7=strongly like)

Contrasts:

1. <u>Timing-only vs. Gain-only:</u>
   Compare mean change from baseline (0,0) for timing-only conditions vs. gain-only conditions.
2. <u>Combined vs. single-factor:</u>
   Compare simultaneous combined variation- (Time, Gain) as (Low, Low), (Med, Med), (High, High) against corresponding single-factor levels, e.g., low, medium, and high levels of time jitter alone or gain variation alone

## Results

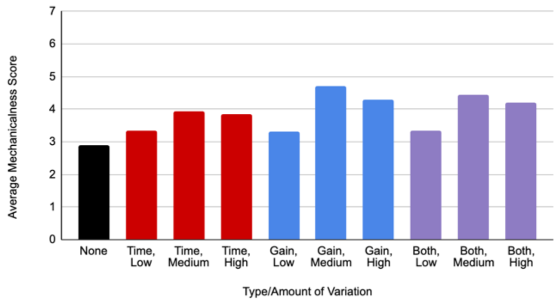


*Figure 3. Mean Mechanicalness Score by Variation Type and Amount*

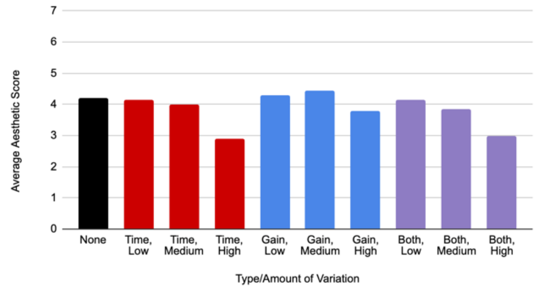


*Figure 4. Mean Aesthetic Score by Variation Type and Amount*

Participants rated two grooves (Patterns A and B) under 10 humanization conditions on a 7-point scale for mechanicalness (1 = totally mechanical, 7 = totally human/natural) and aesthetic preference (1 = strongly dislike, 7 = strongly like). Table values below report the means collapsed across the two patterns.

*Table 2. Mean Mechanicalness Score by Variation Type and Amount for Patterns A and B*

| **Variation Type** | **Pattern A** | **Pattern B** | **Mean Mechanicalness Score (1-7)** |
|---|---|---|---|
| Baseline (None) | 3.00 | 2.80 | 2.90 |
| Time - Low | 3.20 | 3.50 | 3.35 |
| Time - Medium | 3.70 | 4.20 | 3.95 |
| Time - High | 3.10 | 4.60 | 3.85 |
| Gain - Low | 3.30 | 3.30 | 3.30 |
| Gain - Medium | 4.80 | 4.60 | 4.70 |
| Gain - High | 3.70 | 4.90 | 4.30 |
| Both - Low | 3.10 | 3.60 | 3.35 |
| Both - Medium | 4.30 | 4.60 | 4.45 |
| Both - High | 4.20 | 4.20 | 4.20 |

As shown in Table 2, humanness increased (score increased, meaning mechanicalness decreased) relative to baseline in all humanized conditions. The largest humanization increase occurred for Gain-Medium (score = 4.70) followed by Both-Medium (score = 4.45) and Gain-High (score = 4.30).

*Table 3. Mean Aesthetic Score by Variation Type and Amount for Patterns A and B*

| **Variation Type** | **Pattern A** | **Pattern B** | **Mean Aesthetic Score (1-7)** |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Baseline (None) | 3.90 | 4.50 | 4.20 |
| Time - Low | 4.00 | 4.30 | 4.15 |
| Time - Medium | 3.30 | 4.70 | 4.00 |
| Time - High | 2.70 | 3.10 | 2.90 |
| Gain - Low | 3.70 | 4.90 | 4.30 |
| Gain - Medium | 4.20 | 4.70 | 4.45 |
| Gain - High | 3.30 | 4.30 | 3.80 |
| Both - Low | 4.00 | 4.30 | 4.15 |
| Both - Medium | 3.40 | 4.30 | 3.85 |
| Both - High | 2.60 | 3.40 | 3.00 |

Aesthetic preference showed that increased jitter led to lower scores overall, but the opposite was true for gain variation, with the highest score found in Gain-Medium (score = 4.45). Time variation scores, with lowest scores found in Time-High (score = 2.90) and Both-High (3.00). High timing jitter reduced preference strongly, and combined manipulations did not outperform gain-only for preference (Both-Medium score = 3.85, Both-High score = 3.00). Gain variations displayed a slight increase over the baseline.

<u>H1 (Naturalness / mechanicalness)</u>
The hypothesis that timing variation would lead to better humanization (higher scores) than gain variation was not supported, since higher humanization scores were reported for gain-modulated conditions than time-modulated conditions.
<u>H2 (Aesthetic preference):</u>
The hypothesis that gain variation would lead to higher aesthetic scores than timing variation was supported, since the highest aesthetic scores were reported for grooves with only gain modulation applied.
<u>H3 (Interaction / combined humanization):</u>
The hypothesis that for both mechanicalness and aesthetic scores combined gain/time variation would outperform any variation alone was not supported. Gain variation alone, at every variation level (low, medium, and high), with the exception of low variation, outperformed combined gain/time variation.
<u>H4 (Optimal level):</u>
The hypothesis that medium levels of variation would yield the higher scores than the highest level of variation (suggesting an optimal level before humanization is too extreme) was partially supported. For mechanicalness scores, every condition supports this hypothesis: for both time, gain, and combined variation, the medium level showed the highest score, followed by high, then low. For aesthetic scores, this gain variation supported the hypothesis by showing the highest score in the medium level, followed by low, and then high, but the other variation types (time and combined variation) showed consistently lower scores as variation increased (highest score for low, lowest score for high).

**Discussion**

Across one-shot sample-based TR-707 drum grooves, adding variation generally increased perceived humanness (reducing mechanicalness), but the type and amount of variation mattered. Gain variation—especially at a medium level—produced the largest increase in perceived humanness and the highest aesthetic ratings, whereas high timing jitter substantially reduced aesthetic preference despite often increasing perceived humanness relative to the baseline. Combined timing and gain did not consistently outperform gain-only and, for preference, tended to be worse than gain-only at medium/high levels. Overall, the results suggest that for repeated electronic one-shots, loudness micro-variation may be a more effective and safer "humanization" cue than temporal jitter, with an apparent "too much variation" region that reduces liking.

Prior work (e.g., Danielsen et al., 2024; Hennig et al. 2011) emphasizes microtiming as central to groove and perceived naturalness, and many producers presume timing jitter is the main fix for quantized "machine-like" rhythms. This is the main reasoning for DAWs to include a "humanize" button as previously described. However, in these results gain variation produced equal or larger increases in humanness ratings rather than timing jitter, particularly at medium and high amounts. One interpretation is that for one-shot samples that are not diverse in timbre, small loudness fluctuations may function as a proxy for the expressive/timbral variability that occurs naturally in acoustic human performance. Timing jitter alone may not supply enough humanization when the sample itself is identical hit to hit; alternatively, timing offsets may be perceptually interpreted as performance errors when the sound is otherwise completely identical.

The aesthetic preference results align with the idea that some deviation is enjoyable but too much is detrimental. High timing jitter produced a large drop in liking, consistent with the H4 prediction that large timing deviations cross a threshold from "natural" to "sloppy". This resonates with findings that listeners can detect and utilize microtiming statistics (Kaplan et al., 2023) and may dislike deviations that violate stylistic expectations. It also matches the practical producer intuition that humanization should be subtle. Importantly, the participant pool may have elevated sensitivity to timing irregularities, since many subjects reported having music production experience with DAWs as part of their background. This could make high jitter particularly aversive.

In terms of combined variation (both timing and gain simultaneously), humanization does not seem to be additive for preference ratings: adding timing jitter on top of gain variation often reduced liking compared to gain-only. This suggest that timing jitter may introduce perceived instability that outweighs any humanness benefit, particularly for electronic drum-machine timbres. Another possibility is that applying jitter independently to each event (including hi-hats) can create a "smeared" texture; producers often humanize selectively (e.g., less or no variation on hi-hats, or correlated offsets across different pieces of the kit) which was not modeled by the manipulations in this study.

A useful source for interpreting the present aesthetic preference results comes from a 2012 study by Janata et al. which argues that groove reflects a pleasurable drive to move that arises when listeners can couple their motor system to the temporal structure of the music. In other words, listeners prefer grooves that are "danceable". In their experiments, higher-groove stimuli elicited more movement, while stimuli with increased coupling difficulty reduced the feeling of "being in the groove" (Janata et al., 2012). While the present study did not measure participant movement or tapping, the sharp drop in aesthetic preference at high timing jitter levels is consistent with the idea that too much onset variability (timing jitter) reduces temporal predictability (the "feeling" of the groove) and makes entrainment harder, reducing pleasure. By contrast, moderate gain variation may enhance the liveliness of the groove, varying the accent structure without destabilizing timing. This could lead to better entrainment, allowing listeners to feel more "in touch" with the groove. This may explain why gain variation produced the most robust preference benefit.

Strengths

This study was motivated by the use of one-shot samples commonly incorporated in electronic drum grooves to effectively address problems faced by music producers. The use of TR-707 and simple, synthesized grooves closely matches common production workflows and directly targets the "machine gun effect." Additionally, the Gaussian approach provides a simple model for humanization amounts which expands upon the common "Humanize" operation provided by many DAWs.

Limitations

There are several points of consideration to be made while interpreting the data from this study. First, most trials were conducted with closed-back headphones in an environment with a non-negligible level of background noise. Nearly all trials were conducted in a classroom with many individuals present who were talking and listening to music, so differences in background noise and levels of attention for the subjects in this study may inflate variance and reduce statistical power. Additionally, the WAV audio clips (stimuli) were normalized to prevent clipping. While this is useful for managing distortion, this process may have unintentionally compressed the intended gain-variation differences across conditions. This could blur or shift aesthetic preference effects for conditions which included gain variation. Also, pattern A vs. B differed in syncopation (different kick drum patterns), and it is possible that some listeners may have systematically preferred one pattern regardless of humanization. Finally, timing variation was applied based on raw values in milliseconds. For a more accurate representation and variation approach more aligned with previous studies (e.g., Danielsen et al., 2024), it may have been more effective to express and apply timing variation as percentages of the groove tempo rather than absolute values in milliseconds.

Future Work

Future studies could explore more complicated patterns for humanization. Particularly, based on the results and discussion above, it is reasonable to assume that variation applied independently between different elements of the drum kit (different amounts for hi-hat, snare, and kick) could yield more ideal results for both decreased mechanicalness and aesthetic preference. There is also an interesting opportunity to apply more advanced timbral modulations rather than gain changes to more accurately capture human drum sound variation, such as filtering, envelope differences (changing attack and decay), or even alternating between different one-shot sounds for the same drum kit element (i.e. rotating between three snare sounds throughout the groove, or switching randomly between them).

It would also be useful to broaden the sampling of participants in order to compare expert vs. non-expert listeners, i.e. those with experience in music performance/production and those without. It may be possible that a participant's amount or nature of musical training is tied to their level of tolerance for timing jitter.

Additionally, it is reasonable to assume that expanding the study beyond a single drum machine profile (TR-707) and two simple grooves could yield more comprehensive or generalizable results. It could be possible to create stimuli using a wider variety of sounds and rhythms to measure the effectiveness of humanization for different genres of drum grooves (jazz, electronic, pop, etc.)

## Conclusion

This study tested whether two common "humanization" processes for one-shot drum samples—microtiming jitter and hit-to-hit gain variation—can reduce perceived mechanicalness and improve aesthetic preference in electronic drum grooves that otherwise exhibit the "machine gun effect." Across two simple 4-bar patterns, adding either timing or gain variation generally increased perceived humanness relative to a fully quantized baseline, but the type and amount of variation mattered. Contrary to the prediction that timing would dominate naturalness judgements, moderate gain variation produced the largest increase in perceived humanness and also yielded the highest aesthetic ratings, while high timing jitter sharply reduced liking. Combined gain and timing manipulations did not reliably outperform gain-only conditions, suggesting that basic humanization is not necessarily additive when applied as independent Gaussian noise to every event. This experiment contributes a targeted test of humanization in a realistic music production scenario and motivates future work that uses larger and more diverse participant samples, broader sound sets and rhythmic patterns, and more performance-relative models to better determine whether certain humanization techniques enhance both naturalness and musical enjoyment and could provide practical guidance for music producers, helping them choose humanization settings to achieve less mechanical, more appealing drum tracks with minimal editing time.

## References

Ayyildiz, C., Milne, A. J., Irish, M., & Herff, S. A. (2025). Micro-variations in timing and loudness affect music-evoked mental imagery. *Scientific Reports, 15*(1), 30967. https://doi.org/10.1038/s41598-025-12604-4

Danielsen, A., Brøvig, R., Bøhler, K. K., Câmara, G. S., Haugen, M. R., Jacobsen, E., Johansson, M. S., Lartillot, O., Nymoen, K., Oddekalv, K. A., Sandvik, B., Sioros, G., & London, J. (2024). There's more to timing than time: Investigating musical microrhythm across disciplines and cultures. *Music Perception, 41*(3), 176–198. https://doi.org/10.1525/mp.2024.41.3.176

Hennig, H., Fleischmann, R., Fredebohm, A., Hagmayer, Y., Nagler, J., & Geisel, T. (2011). The nature and perception of fluctuations in human musical rhythms. *PLOS ONE, 6*(10), e26457. https://doi.org/10.1371/journal.pone.0026457

Jakubowski, K., Polak, R., Rocamora, M., Jure, L., & Jacoby, N. (2022). Aesthetics of musical timing: Culture and expertise affect preferences for isochrony but not synchrony. *Cognition, 227*, 105205. https://doi.org/10.1016/j.cognition.2022.105205

Janata, P., Tomic, S. T., & Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. Journal of experimental psychology. General, 141(1), 54–75. https://doi.org/10.1037/a0024208

Kaplan, T., Jamone, L., & Pearce, M. (2023). Probabilistic modelling of microtiming perception.
*Cognition, 239*, 105532. https://doi.org/10.1016/j.cognition.2023.105532