Bachelor thesis

Artificial Intelligence

Radboud University

Predictability and Uncertainty in the Pleasure of Music using the Music Transformer

Author:
Pien Rooijendijk
s1054190

First supervisor:
dr. Eelke Spaak
Donders Centre for Cognition
eelke.spaak@donders.ru.nl

Second reader:
dr. Kiki van der Heijden
Donders Centre for Cognition
kiki.vanderheijden@donders.ru.nl

September 3, 2025

Abstract

Pleasure of music in humans arises from the interaction between the reward system and the auditory areas of the brain; humans find music rewarding. Previous research by Gold et al. (2019) already showed significant results by using the Information Dynamics of Music Model (IDyOM) to measure predictability and uncertainty in music. This thesis will measure these same factors of music by using the Music Transformer (MT) to explain the same data. It can then be possibly concluded that it is preferable to use the MT rather than the IDyOM.

Contents

Chapter 1

Introduction

One of the greatest pleasures among humans is music. This pleasure arises from the interaction between the reward system and the auditory areas of the brain; humans find music rewarding (Gold et al. 2019). Dopamine cells encode to what extent the outcome matches our expectations, there are stronger responses to outcomes that are better than expected (Salimpoor et al., 2015). Expectations of the sequence of different notes in music shape our experience of music, the prediction error from the dopamine cells helps us to improve future predictions. According to Sloboda (1991), emotion is based on confirmations and expectancy violations. Music is often most pleasurable when these changes in notes are dramatic and sudden. These changes are liked by humans when they are in naturalistic and familiar music (Sloboda, 1991). Yet these surprises are unpleasant when the context is lacking, or when the expectations are violated.

There are two aspects of expectations in music according to Salimpoor et al. (2015): the knowledge of how a musical piece, which is familiar, will unfold, and the implicit understanding of the rules of music based on the music-listening history someone has. Social and cultural influences can affect these aspects by using human statistical learning. But how can a piece of music be better than we expected? According to Salimpoor et al. (2015) the answer lies in the sheer complexity of music. This includes that a musical piece needs to be complex; it has to have changing harmonic, unique expressive features of a performer's personal style and spectral and rhythmic features (Salimpoor et al., 2015). When listening to musical pieces with these properties, we have expectations of which event will be next. When to expect an event, relates to matching the structure of the musical piece with the rhythm of the music which then can be extrapolated in the future (Rohrmeier & Koelsch, 2012). Could the expectancy (entropy) and melodic surprises (information content) of notes help explain the pleasure of music?

Gold et al. (2019) already studied two key aspects of musical complexity, predictability and uncertainty. Two studies have been conducted by Gold et al. (2019) where they evaluated how uncertainty and predictability affect musical preferences in human participants, using the Information Dynamics of Music Model (IDyOM). Gold et al. (2019) found the Wundt effect in their results; the effect which links pleasure to intermediate levels of arousal (Wundt, 1948). This phenomenon shows a U-shaped curve in the results where its effect is pleasant or rewarding. The peak of its curve is the preferred level of predictability (Lisøy et al., 2022). However, when the values are increased to higher levels, it can be experienced as unpleasant. In the results of both studies, the Wundt effect can be found between the liking ratings and both uncertainty and entropy. In this Bachelor Thesis, the predictability and uncertainty of music will be studied, using a transformer neural network: the Music Transformer (MT). Since the state-of-the-art model can cap-

ture long-range dependencies and can be trained on polyphonic music, can it also evaluate the musical preferences in the same human participants from the research from Gold et al (2019) better than the Information Dynamics of Music Model did? Also, can the Wundt effect in the results of the probability distribution from the MT be found again using the same human liking ratings?

1.1 The Wundt effect

It is probable that humans will dislike music that surpasses their level of comprehension or processing capability. It is also feasible that music which is significantly less complex than an individual's processing ability may be perceived as dull, repetitive, and tedious, leading to a dislike of such music (Madison & Schiölde, 2017). The Wundt curve (Wundt, 1874), captures these two aspects. Gold et al. (2019) tested for the Wundt effect between complexity and liking (see figure 1.1). Berlyne (1971) already linked the Wundt effect with pleasure to intermediate levels of arousal. Berlyne's experiments and his successors have focused on examining the impact of collative variables on preference. Among these variables, complexity and familiarity have received significant attention (Chmiel & Schubert, 2017). Another research by Rossing & Stumpf (1998) named "The Science of Sound" also discusses the Wundt curve in the context of musical perception. The authors explain how the Wundt curve can be used to enhance musical expression and create a more dynamic musical experience for the listener.

What Gold et al. (2019) concluded from their results is that the participants preferred music with medium complexity more than simple and highly complex music. This conclusion is also supported by Madison & Schiölde, (2017), who found that complex music requires more hours to listen to in order for the listener to develop some level of familiarity that is required to get pleasure from it. The same liking ratings from the participants will be used to compare the information content (IC) and entropy from the MT to search for the Wundt effect. If the Wundt effect is found in the results, the preferences of the human participants for the different degrees of complexity of music can be discovered.

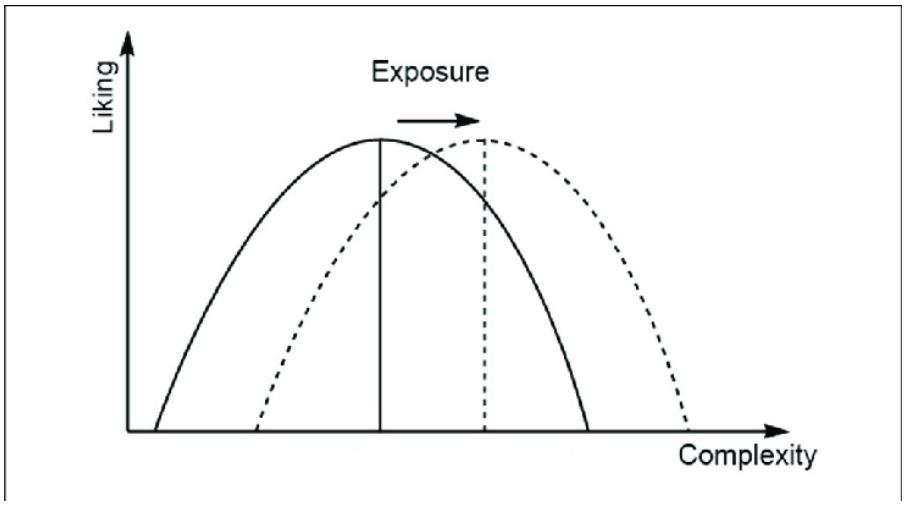


Figure 1.1: The Wundt curve (Madison & Schiölde, 2017)

1.2 Why use the Music Transformer?

The Music Transformer and the Information Dynamics of Music (IDyOM) model are both powerful models for generating music. While they share some similarities, there are several advantages to using the Music Transformer over the IDyOM model:

1. **Better long-term dependencies**: In the IDyOM model, the Markov assumption limits the ability of the model to capture long-term dependencies. The Markov property in the IDyOM restricts the model's consideration of dependencies between musical events to a fixed number of preceding events, it is a simplifying assumption (Pearce, 2018). The length of how many preceding events the IDyOM is considering is determined by the order of the Markov assumption. The Music Transformer, on the other hand, uses self-attention mechanisms (see section 1.3) to better capture long-term dependencies. It also captures relationships between the different events in the entire input sequence (Huang et al., 2018). So while the IDyOM has local coherence, it may limit the model's ability to capture long-term dependencies and generate music with more global coherence compared to the MT.

2. **Increased flexibility**: The Music Transformer is also more flexible than the IDyOM model, as it can also be trained on polyphonic music and different styles of music (Huang et al., 2018). The IDyOM model, on the other hand, is more specialized and designed specifically for Western classical music (Pearce & Wiggins, 2004). Furthermore, the MT can be fine-tuned for a wide variety of tasks, including music composition arrangement, and transcription, whereas the IDyOM model is primarily designed for music prediction.

3. **Better performance**: Kern et al. (2022) tested the performances of the MT and IDyOM models with which these models predicting upcoming notes, given prior notes. The MT performed slightly better than the IDyOM, the IDyOM had an accuracy level of 53.5% and the MT 54.8%. Additionally, Kern et al. (2022) found this increased predictive accuracy score for the MT when the context length k increased, from 9.17% for k = 1 up to 54.82% for k = 350. The IDyOM, however, had an optimal performance when the context had a length of 4, this was caused by the Markov assumption. In terms of note-level surprise, the MT also performed better. The predictive performance measurement for both models involved assessing the median surprise across compositions. Lower values indicated a greater ability to predict the next note based on the context of the current note. The IDyOM scored 1.46 and the MT 1.15 (Kern et al., 2022).

While both models are suited for generating and analysing music, the MT has significantly more advantages in terms of its flexibility, performance and its ability to capture long-term dependencies.

1.3 The internals of the Music Transformer

The Music Transformer is a Transformer model with self-attention, meaning it can model long-range dependencies which makes it robust (Tay, 2020). It has access to the output which was generated previously and computes the next steps efficiently. Neural networks have to store this memory which they want to access later in a fixed-size memory state. Saving all these previous states can make training way more difficult (Huang, 2018).

According to Huang (2018), having these self-attention mechanisms, the MT can capture the self-referential characteristics music has. Transformer models process all notes of music as a whole, whereas other neural networks take sequences of variable length as input.

1.3.1 Data Input

For the input for training the MT, a language-modelling approach is used. The music is encoded as a sequence of tokens, and the different notes the MT knows will be determined by the training dataset (Huang, 2018). The total amount of MIDI pitches is 128, which indicates the onset of MIDI pitches from 0 (C-1) to 127 (G9) (Y. Huang & Yang, 2020). The input sequence is encoded according to different events, see table 1.1 for the definitions.

Event	Definition
128 `NOTE_ON` events	Starting a note with one of the 128 MIDI pitches.
128 `NOTE_OFF` events	Ending a note with one of the 128 MIDI pitches.
100 `TIME_SHIFT` events	Relative time gap between events.
32 `SET_VELOCITY` events	Velocity of coming `NOTE_ON` events, where the 128 possible MIDI velocities are quantized into 32 bins.

Table 1.1: Table to test captions and labels (Huang, 2018).

The tokens are derived from a symbolic representation of the music, which are MIDI files in this research. This representation has its roots in the MIDI files since these are also using `NOTE_ON` and `NOTE_OFF` events (Y. Huang & Yang, 2020).

1.3.2 Transformer Encoder

The encoder consists of multiple layers of feed-forward neural and self-attention networks (Zhang, 2021). Each layer consists of two sublayers. The first sublayer is a multi-head self-attention mechanism, while the second sublayer is a feed-forward network. The encoder eventually gives a vector representation as output for each position of the input sequence.

1.3.3 Transformer Decoder

The decoder of the MT is programmed with multiple identical layers as the encoder. Every layer in the decoder has three sub-layers which form a residual connection followed by layer normalization (Zhang, 2021). It can, with the help of attention mechanisms, review an entire sequence and it can choose the different notes to decode (Rahali & Akhloufi, 2023). When generating a musical piece, given a prior, the final layer of the decoder generates a probability distribution of the whole vocabulary of possible tokens on which the model was trained on. This final layer consists of a linear transformation followed by a softmax activation function (Huang, 2018).

Chapter 2

Methods

Gold et al. (2019) found the Wundt effect in their data by comparing the IDyOM's probability distributions of the IC and entropy with the liking ratings of the participants. In the section below (section 2.1), a brief description is given of how Gold et al. (2019) required their human data, and how this data will be used to compare it to the IC and entropy computed by the MT (section 2.4). With this comparison based on the probability distributions from the MT, the liking rating data will be plotted with both the IC and entropy. Analysing these plots will show whether there is a Wundt effect to be found, and whether this effect is more present when using the MT rather than the IDyOM model. Only the data from the first study by Gold et al. (2019) were used.

2.1 Participants and procedure

In this research, the liking ratings from the human participants from the study by Gold et al. (2019) were used. Their study involved 44 healthy volunteers (25 females, mean age $\pm$ SD = 21.56 $\pm$ 3.31 years) with normal hearing who listened to the 55 musical excerpts and rated their level of liking for each musical piece on a scale ranging from 1 (very little) to 7 (very much). Before the listening task, participants completed three questionnaires on their musical sophistication, the degree to which they associate music with reward, and their personality traits. In addition to the rating task, the participants were assigned an orthogonal task of pressing the "Enter" key as soon as they detected a timbre change in the stimulus. This secondary task aimed to ensure their attentiveness during the listening task. Participants were asked to exclude any excerpts they recognized to avoid a possible relationship between familiarity and musical predictability. From the 2337 trials in total, 431 trials were rated as familiar by the participants, leaving 1906 stimuli in total for the analysis.

There were four alternative viewpoints suggested of the IDyOM by Gold et al. (2019). A total of seven different configurations were modeled, and the selection of these models was based on a comparison between the IC output of each model and the unexpectedness ratings provided by a separate sample of 24 participants. This second sample consisted of 17 females and 7 males, with a mean age of 22.08 years ($\pm$ 2.70 standard deviation) and a mean musical experience of 2.89 years ($\pm$ 4.52 standard deviation) [1]. Subsequently, the musical pieces were sorted into five clusters of mean duration-weighted information content (mDW-IC), which were computed in the selection trial, and presented to the participants in a random and participant-specific order. Two participants were excluded

[1] It's important to note that these participants did not take part in the initial study.

from the analysis. One rated every stimulus as familiar. Another withdrew from the experiment halfway through, and the data which was already collected were used.

For more details, see section Study 1, Materials and Methods in the research paper by Gold et al. (2019).

2.2 Stimuli

The 55 stimuli were snippets from real composed music pieces retrieved from the public Musical Instrument Digital Interface databases. The following websites were used for the retrieval: www.osk.3web.ne.jp/∼kasumitu/eng.htm and www.classicalarchives.com/midi.html. The MT was used to compute the unpredictability and uncertainty of all 55 stimuli. Monophonic stimuli were used to avoid the effects of harmony and polyphony, the peak amplitudes were also normalized to the same level, and the tempo was also changed to either 96, 120 or 144 bpm. The stimuli were transformed into WAV files that possess a naturalistic quality. The conversion process was carried out using the Kontakt 5 synthesizer developed by Native Instruments (2018). The execution of the conversion took place within the Ableton Live 9 digital audio workstation, developed by Ableton (2018). Each musical piece was generated using a digital flute synthesizer, except for the attention trial stimuli. Digitial filtering techniques were applied to simulate the more natural sound of a music studio. To enhance the organic feeling of the stimuli and prevent them from sounding mechanical or artificial, Gold et al. (2019) introduced slight random shifts to the note onsets at a millisecond scale using Ableton's Groove Pool with a 25% randomization factor, thereby incorporating a more human touch to the stimuli.

For more specifications about the stimuli, see appendix A.

2.3 Computational modeling

2.3.1 Training

The MT is trained by Kern et al. (2022) on the polyphonic Maestro corpus, which can be found on: https://magenta.tensorflow.org/datasets/maestro. This dataset contains 200 hours of audio and MIDI recordings performed by various artists. My script is based on the open adaptation for PyTorch https://github.com/gwinndr/MusicTransformer-Pytorch and code from Kern et al. which is publicly available on the Donders Repository https://data.donders.ru.nl/collections/di/dccn/DSC_3018045.02_116?0. Although Gold et al. (2019) used their own corpus of Western music, I will be using the pre-trained MT. Kern et al. (2022) used 300 epochs and the training parameters from the original paper by Huang et al. (2018) (learning rate = 0.1, batch size = 2, number of layers = 6, number of attention heads = 6, dropout rate = 0.1). The progress of the training based on the cross-entropy loss computed on both the training and test data (80%-20%) was monitored. The cross-entropy loss, which represents the average surprise across all notes, served as the metric to train the model in minimizing the surprise associated with upcoming notes. They achieved a minimum loss of 1.97, and this was comparable to the value reported in the paper by Huang et al. (2018) which was 1.835. After the initial training, the model was finetuned to adjust to monophonic music. The same training parameters were used when training on the Monophonic Corpus of Complete Compositions (MCCC) dataset (https://osf.io/dg7ms/). The weights which were used for the pre-trained MT are obtained by Kern et al. (2022) at epoch 21.

2.3.2 Computations

Using the pre-trained model, it computes the probability distribution X_t of the next note at time point t given k preceding consecutive notes using a sequence-to-sequence method:

$$P(X_t|x_{t-k}^{t-1}), \text{where } X \in 0...127, k > 0, t \geq 0 \tag{2.1}$$

The first note in the composition has a uniform distribution for each possible note, which will be $P(X_0 = x) = 1/128$. x stands for a continuation or note pitch from the vocabulary X.

Each song thus had for each note a probability given the prior note. The IC and entropy of all 55 stimuli were computed by using this probability distribution $P(X_t|x_{t-k}^{t-1})$. The following two formulas were used to compute the IC and entropy:

$$\text{IC}(x_t) = -\log_e(P(x_t|x_{t-k}^{t-1})) \tag{2.2}$$

$$\text{Entropy}_t = -\sum_{x=0}^{127} P(X_t = x_t|x_{t-k}^{t-1}) \times \log_e P(X_t = x_t|x_{t-k}^{t-1}) \tag{2.3}$$

In this context, IC refers to information content, which serves as a representation of the surprise or unexpectedness of a musical piece. On the other hand, entropy is used to represent the level of uncertainty or instability within the music. Both formulas are used to compute all possible continuations X, where x is each possible continuation from an alphabet X. This alphabet X will consist of the pitch values of the notes ranging discretely from 0 to 127. Other values such as time shift and velocity values are disregarded when computing the IC and entropy.

Gold et al. (2019) computed the IC and entropy just as above, but they did not treat all events equally. Considering the duration of all events and assigning higher weights to longer events, they ensured that each 30-second stimulus is represented as one unit. Events which have a longer duration, contribute more significantly to the overall measure. The resulting mean duration-weighted information content (mDW-IC) and mean duration-weighted entropy (mDW-Ent) provide a more comprehensive representation of the impact of the different notes, considering both their informational content and how long they persist.

2.4 Experimental design and statistical analysis

Excluding participant 15, which rated every stimulus as familiar, the remaining trials were tested for linear and quadratic effects. Using the `fitlm` function in MATLAB, linear and quadratic regression were performed using the mDW-IC and mDW-Ent as predictor variables. Both variables predicted the liking ratings separately of all the stimuli from the participants. Gold et al. (2019) originally chose a linear mixed-effect model, but for an optimal comparison, the results were also estimated by linear and quadratic regression. The linear and quadratic effects were evaluated using MATLAB functions. ANOVA measurements were used to compute the variance of the MT based on the cluster.

I am researching how musical surprise might be involved with the uncertainty of a musical note, which could affect the liking ratings of the participants. To avoid possible collinearity of mDW-Ent and mDW-IC, each stimulus is classified according to its mDW-IC and mDW-Ent into three groups using the k-means clustering algorithm from MATLAB. Without using the participants liking ratings, the algorithm identified three

groups with Euclidean distance minimization. The category of low mDW-IC and low mDW-Ent consisted of 20 stimuli. There were 12 stimuli classified as having medium mDW-IC and medium mDW-Ent, while 23 stimuli were categorized as having high mDW-IC and high mDW-Ent (see figure 2.1). These groups represent a robust classification of the stimuli based on the mDW-Ent and mDW-IC. This clustering was used to test again for the Wundt effect using the ANOVA model. The different categories, medium, high and low mDW-IC and mDW-Ent, were compared for researching the preference of the participants for the complexity of the 55 musical pieces. The interaction between the different degrees of musical complexity was investigated with the post hoc Tukey-Kramer Significant Difference test.

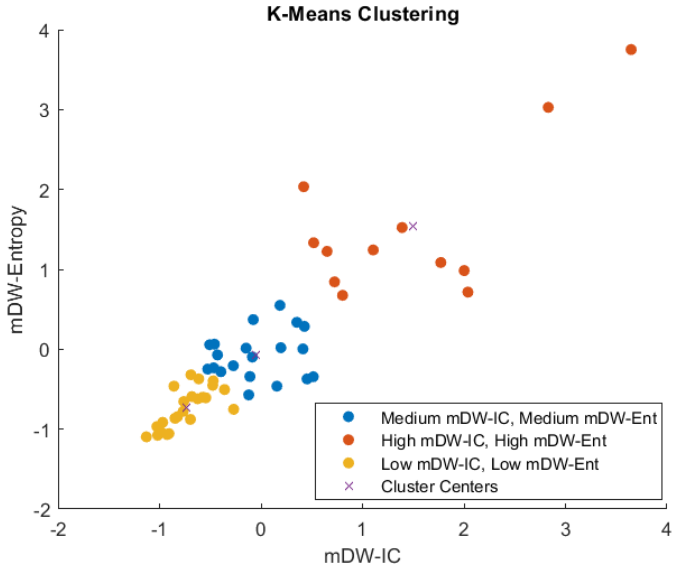
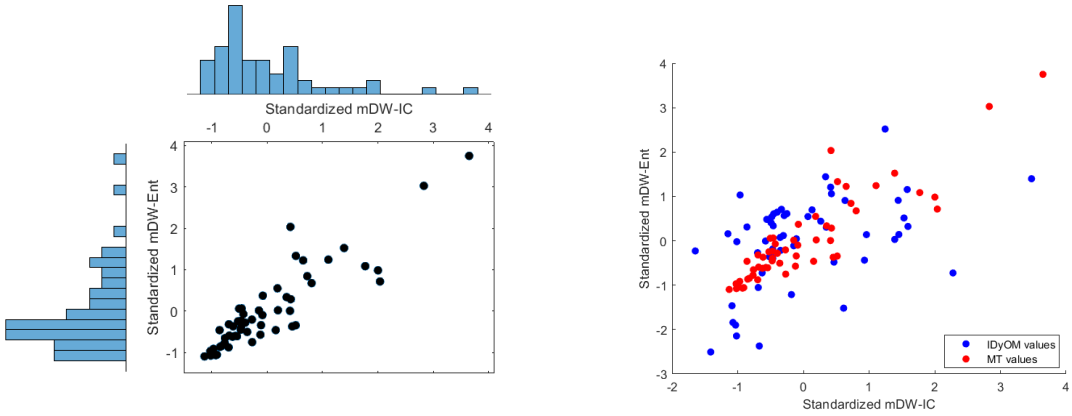


Figure 2.1: k-means clustering of the 55 stimuli

Chapter 3

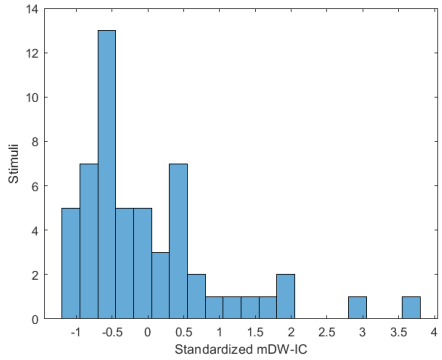
Results

mDW-IC and mDW-Ent measured by the MT were found to be highly positively correlated (Pearson's $r = 0.8972$, $p < 0.001$, see also figures 3.1a, 3.2a and 3.2b). These results verify that when the IC of one of the notes increases, the overall entropy of the distribution also increases, indicating higher uncertainty. They are correlated since IC and entropy both are related concepts which describe the uncertainty of the musical pieces.

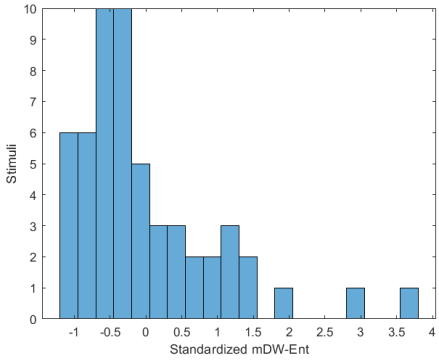


(a) (b)

Figure 3.1: **(a)** Stimulus unpredictability and uncertainty distributions from the MT. **(b)** Stimulus unpredictability and uncertainty distribution for both the IDyOM and MT.



(a) Standardized mDW-IC for all stimuli



(b) Standardized mDW-Ent for all stimuli

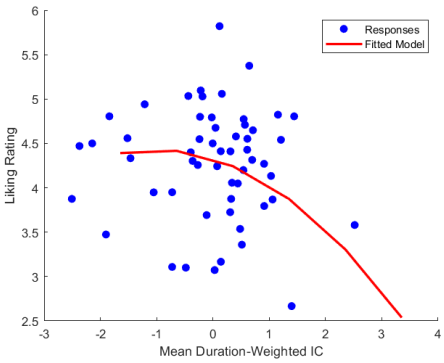
Figure 3.2: The standardized metrics for all stimuli from the MT

3.1 Models and regression

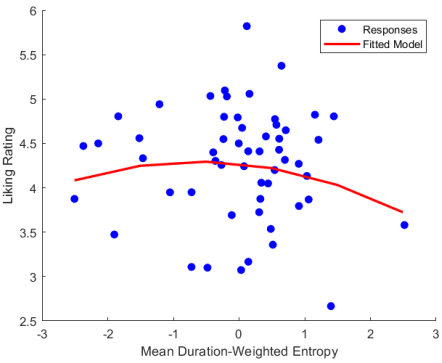
3.1.1 Measurements from the IDyOM

Gold et al. (2019) found a significant Wundt effect in their results for both mDW-IC and mDW-Ent which is also modelled in figures 3.3a and 3.3b by doing linear and quadratic regression again on the human data and output of the IDyOM model [1]. The optimal model for mDW-IC showed a significant negative linear effect ($\beta = -0.20, p < 0.001$). Additionally, the model also exhibited a significant negative quadratic effect, where $\beta = -0.10, p < 0.001$. Using ANOVA measurement, the model explained 22,6 % of the variance in liking ratings ($p < 0.001$).

Between the mDW-Ent and liking ratings was also a Wundt effect present, the optimal model contained a significant negative linear effect where $\beta = -0.07, p = 0.045$ and a negative quadratic effect where $\beta = -0.06, p = 0.013$. Using ANOVA measurement, the model explained 4,0 % of the variance in liking ratings ($p = 0.049$).



(a)

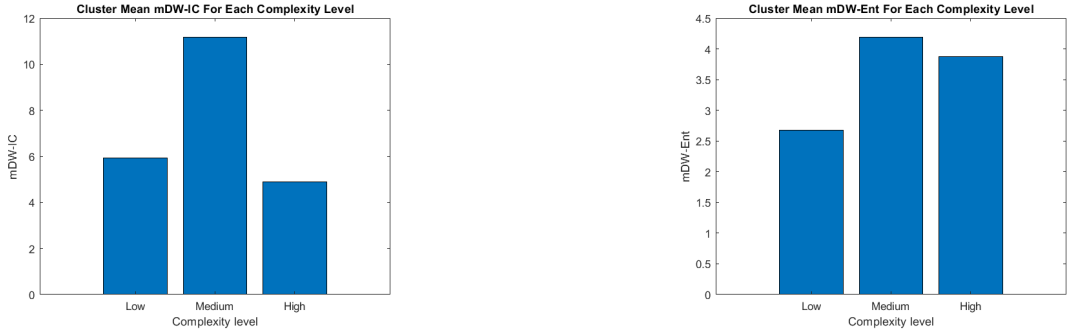


(b)

Figure 3.3: **(a)** The optimal model from the IDyOM of mDW-IC by Gold et al. (2019). **(b)** The optimal model from the IDyOM of mDW-Ent by Gold et al. (2019).
The red curve is the fitted quadratic model and the blue dots are representing the mean liking ratings for each of the 55 stimuli.

[1] Note that these plots, constructed by using quadratic regression, are similar to the original ones from the research paper by Gold et al. (2019), which used a linear mixed-effect model.

Another k-means clustering was done on the IDyOM values from Gold et al. (2019) where there were only 3 clusters involved. Originally, there were 6 clusters. The ANOVA measurement was used to further investigate the preferences for the different levels mDW-IC and mDW-Ent (see figures 3.4a and 3.4b). Both mDW-IC and mDW-Ent have an effect on the liking ratings ($p < 0.001$) (see figures 3.5a and 3.5b). Using the post hoc Tukey-Kramer Honest Significant Difference test again, it can be concluded that for mDW-IC there was a Wundt effect since high mDW-IC < low mDW- IC: $p < 0.001$, high mDW-IC < medium mDW-IC: $p < 0.001$ and low mDW-IC < medium mDW-IC: $p < 0.001$ [2]. Using the same post hoc Tukey-Kramer Honest Significant Difference test it can be concluded that there are also the same preference: low mDW-Ent < high mDW- Ent: $p < 0.001$, high mDW-Ent < medium mDW-Ent: $p < 0.001$ and low mDW-Ent < medium mDW-Ent: $p < 0.001$.



(a) (b)

Figure 3.4: **(a)** ANOVA model for the mean mDW-IC for each complexity level. **(b)** ANOVA model for the mean mDW-Ent for each complexity level.

ANOVA Table

Source	SS	df	MS	F	Prob>F
Groups	287.979	2	143.99	59.25	3.90113e-14
Error	126.37	52	2.43		
Total	414.349	54			

(a)

ANOVA Table

Source	SS	df	MS	F	Prob>F
Groups	13.7463	2	6.87315	56.09	1.0411e-13
Error	6.3717	52	0.12253		
Total	20.1181	54			

(b)

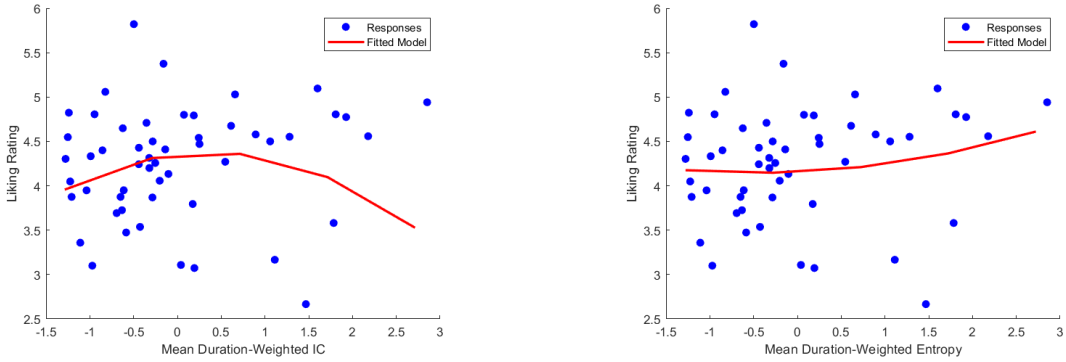
Figure 3.5: **(a)** ANOVA table for the mean mDW-IC for each complexity level. **(b)** ANOVA table for the mean mDW-Ent for each complexity level.

3.1.2 Measurements from the Music Transformer

After testing for outliers using the Z-score method, two stimuli were significant outliers (47_LesFoliesNo5.mid and 48_LeRossignol.mid, see also Appendix A for more details on the stimuli). Excluding these two outliers from the regression modelling (see figures 3.6a and 3.6b), there is a significant Wundt effect between the human liking ratings and mDW-IC computed by the MT. This fitted model contained a significant positive linear effect

[2]Note that for both mDW-IC and mDW-Ent the complexity levels are the same when comparing them. When mDW-IC is low, then mDW-Ent is also low.

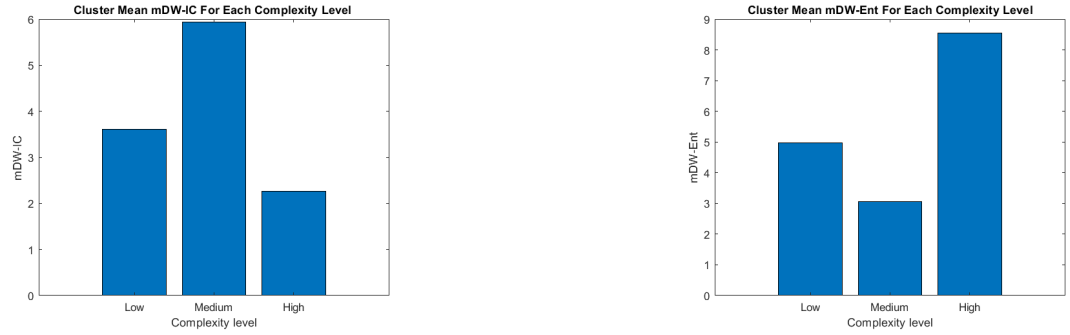
with $\beta = 0.11, p = 0.014$ and a negative quadratic effect where $\beta = -0.15, p < 0.001$. Using the ANOVA measurement, the model explained 6,1 % of the variance in liking ratings ($p = 0.014$). However, there is not a significant Wundt effect between the human liking ratings and mDW-Ent computed by the MT ($p = 0.333$).



(a) (b)

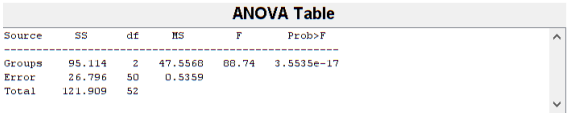
Figure 3.6: **(a)** The model from the MT of mDW-IC. **(b)** The model from the MT of mDW-Ent. The red curve is the fitted quadratic model and the blue dots are representing the mean liking ratings for each of the 55 stimuli.

Using k-means clustering, the 55 stimuli were categorized according to their degree of complexity (see figure 2.1). Then ANOVA was used to further investigate the preferences for the different levels mDW-IC and mDW-Ent (see figures 3.7a and 3.7b). Both mDW-IC and mDW-Ent have an effect on the liking ratings ($p < 0.001$) (see figures 3.8a and 3.8b). Using the post hoc Tukey-Kramer Honest Significant Difference test, it can be concluded that for mDW-IC there was a Wundt effect since high mDW-IC < low mDW-IC: $p < 0.001$, high mDW-IC < medium mDW-IC: $p < 0.001$ and low mDW-IC < medium mDW-IC: $p < 0.001$ (see also figure 3.7a). Using the same post hoc Tukey-Kramer Honest Significant Difference test it can be concluded that there is a significant preference for stimuli with high mDW-Ent ($p < 0.001$).



(a) (b)

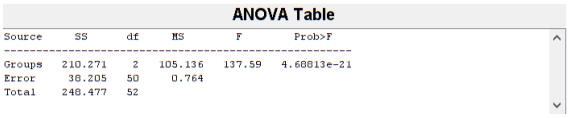
Figure 3.7: **(a)** ANOVA model for the mean mDW-IC for each complexity level. **(b)** ANOVA model for the mean mDW-Ent for each complexity level.



ANOVA Table

Source	SS	df	MS	F	Prob>F
Groups	95.114	2	47.5568	88.74	3.5535e-17
Error	26.796	50	0.5359		
Total	121.909	52			

(a)



ANOVA Table

Source	SS	df	MS	F	Prob>F
Groups	210.271	2	105.136	137.59	4.68813e-21
Error	38.205	50	0.764		
Total	248.477	52			

(b)

Figure 3.8: **(a)** ANOVA table for the mean mDW-IC for each complexity level. **(b)** ANOVA table for the mean mDW-Ent for each complexity level.

As can be seen in figures 3.3a, 3.3b, 3.6a and 3.6b, and also in the ANOVA result plots (see figures 3.4a, 3.4b, 3.7a and 3.7b), compared to the Wundt results form Gold et al. (2019) only the mDW-IC shows a significant Wundt result. The IDyOM explained 22,6% of the variance in liking ratings compared with the mDW-IC and the MT only 6,1% based on the quadratic model. The variance in liking ratings compared with the mDW-Ent was not significant ($p = 0.333$). This analysis implies that the human participants desired more surprising contexts and had a preference for musical pieces with medium mDW-IC.

Chapter 4

Discussion

Using the MT there is a significant Wundt effect to be found between the liking ratings of the participants and mDW-IC. It validates a relation between the complexity of music and how much humans enjoy it. As described above, there is evidence that listeners like medium complexity [1] more than high and low complexity which was measured by the MT. Quadratic models were made based on linear regression for both models, with unpredictability, denoted as mDW-IC and uncertainty, as mDW-Ent as the predictor variables. Comparing the two models used in this research, the IDyOM and MT, they both showed a significant Wundt effect when comparing mDW-IC with the human liking ratings. The IDyOM explained 22,6% of the variance in liking ratings compared with mDW-IC and the MT 6,1% based on the regression analysis. Two outliers were removed from the data to ensure that the significance level was met to get the Wundt effect in the results. Since the explanation of the variance of mDW-Ent compared with the liking ratings was not significant ($p = 0.333$), the MT's measurement could not be compared to the mDW-Ent from the IDyOM. Solely based on this research, I cannot conclude that based on a specific degree of uncertainty, the human participants found the stimuli the most pleasurable. However, given the ANOVA measurement, it showed a significant preference for medium mDW-IC when mDW-Ent was also medium. Although mDW-IC and mDW-Ent are highly correlated (Pearson's $r = 0.8972, p < 0.001$), the ANOVA analysis showed that human preferences are more dynamic than just a preference for higher complexity.

For further research, the optimal values for mDW-IC and mDW-Ent can be taken into account when for example composing new Western tonal music with the MT. These new optimal music pieces can be compared with the training corpus of the model and how these new musical excerpts affect the liking ratings of humans. Considering that humans prefer medium complexity more than high and low complexity, we can incorporate this into the Music Transformer when generating new notes according to the optimal probability distributions. As stated before, it is rewarding to learn about the musical structure of music to make it more pleasurable for humans. It may be rewarding to optimize the entropy and IC, as measured in this research when making new music with the Music Transformer.

The variance which explained the human liking ratings was not expected when using the MT to explain the human data. There are a few factors which could explain this low variance. First, the MT was trained on both the Maestro and MCCC corpus, whereas the IDyOM was trained on a large dataset of Western tonal music. This difference could

[1] Note that complexity here only entails the information content, which are the surprise elements of music.

have affected the alphabet of the MT. The monophonic stimuli were also from the same genre. Since IDyOM is specifically designed for Western classical music, the performance on these Western tonal stimuli was already better (Pearce & Wiggins, 2004). If the MT would have been trained on this same Western tonal corpus, the performance could have been improved. Second, there appears to be a strong positive correlation between the standardized mDW-IC and standardized mDW-Ent (Pearson's $r = 0.8972$, $p < 0.001$, see figure 3.1a). It could have affected the absence of the Wundt effect between the liking ratings and mDW-Ent given the strong positive correlation. This correlation was also slightly present in the results from Kern et al. (2022) when computing the surprise and uncertainty with the pre-trained MT. mDW-IC and mDW-Ent are supposed to be positively correlated since they are both computed from the same probability distribution, but this correlation between the two variables was significantly high compared to the measured mDW-IC and mDW-Ent from the IDyOM (Pearson's $r = 0.44, p < 0.001$). Finally, the interpretation that the participants liked medium complexity more than high and low complexity could have been supported more if polyphonic music stimuli were used instead of monophonic stimuli. Polyphonic music sound more like real-world musical pieces. Polyphonic stimuli were excluded, so the higher complexity distribution was left undersampled.

Since the data for the second study from Gold et al. (2019) was not provided, the additional effects of repeating stimuli have not been researched. If these effects were to be explored, the MT could have possibly shown the Wundt effect again between the liking ratings and mDW-IC. Also, Gold et al. (2019) studied the Wundt effect of each individual participant. Researching the individual Wundt effect would have taught me about the musical sophistication of each participant.

The findings from this thesis are attributed to the computational and predictive power of the MT. The MT is a powerful Transformer network which again found the Wundt effect in the comparison between mDW-IC and human liking ratings. It implies that based on predictions about the surprise and uncertainty of musical notes, we can learn about how music (as complex as it can sometimes be) can be pleasurable to humans.

Chapter 5

Data and source code

The pre-trained Music Transformer from Kern et al. (2022) was used from https://data.donders.ru.nl/collections/di/dccn/DSC_3018045.02_116?0. I adjusted the python file where the MIDI files are processed as a batch to compute the probability distribution of the next node given the preceding notes. These alterations and my own script for computing the Information Content and Entropy can be found on https://github.com/pcrooijendijk/MTBachelorThesis.

Chapter 6

Acknowledgements

A special thank you to Benjamin P. Gold for providing their research results from the first study and Eelke Spaak for providing the code from their Music Transformer model.

Chapter 7

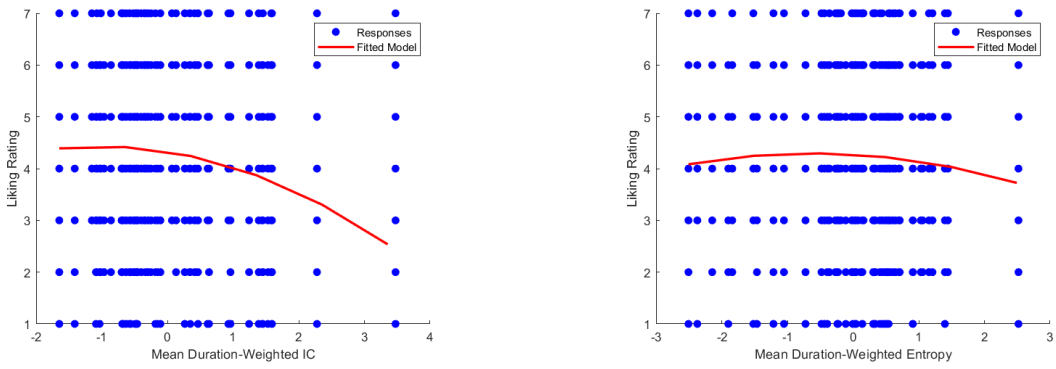
Bibliography

- Berlyne, D.E. (1971) Aesthetics and Psychobiology. Appleton-Century-Crofts, New York. - References - Scientific Research Publishing. (z.d.). https://www.scirp.org/(S(czeh2tfqyw2orz553k1w0r45))/reference/ReferencesPapers.aspx?ReferenceID=2258501
- Chmiel, A., & Schubert, E. (2017). Back to the inverted-U for music preference: A review of the literature. Psychology of Music, 45(6), 886–909. https://doi.org/10.1177/0305735617697507
- Cooper, K. D., & Torczon, L. (2012). Code Shape. In Engineering a Compiler (Second Edition). https://doi.org/10.1016/b978-0-12-088478-0.00007-4
- Gold, B. D., Pearce, M. T., Mas-Herrero, E., Dagher, A., & Zatorre, R. J. (2019). Predictability and Uncertainty in the Pleasure of Music: A Reward for Learning? The Journal of Neuroscience, 39(47), 9397–9409. https://doi.org/10.1523/jneurosci.0428-19.2019
- Gwinn D, Myrick B, Nélias C. 2022. Gwinndr/musictransformer-pytorch. 1.0. Github. https://github.com/gwinndr/MusicTransformer-Pytorch
- Hawthorne, C. (2018, 29 oktober). Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset. arXiv.org. https://arxiv.org/abs/1810.12247
- Huang, C. A. (2018, 12 september). Music Transformer. arXiv.org. https://arxiv.org/abs/1809.04281#
- Huang, Y., & Yang, Y. (2020). Pop Music Transformer. https://doi.org/10.1145/3394171.3413671
- Kern, P., Heilbron, M., De Lange, F. P., & Spaak, E. (2022). Cortical activity during naturalistic music listening reflects short-range predictions based on long-term experience. eLife, 11. https://doi.org/10.7554/elife.80935
- Lisøy, R. S., Pfuhl, G., Sunde, H. F., & Biegler, R. (2022). Sweet spot in music—Is predictability preferred among persons with psychotic-like experiences or autistic traits? PLOS ONE, 17(9), e0275308. https://doi.org/10.1371/journal.pone.0275308
- Madison, G., & Schiölde, G. (2017). Repeated Listening Increases the Liking for Music Regardless of Its Complexity: Implications for the Appreciation and Aesthetics of Music. Frontiers in Neuroscience, 11. https://doi.org/10.3389/fnins.2017.00147
- Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation. Annals of the New York Academy of Sciences, 1423(1), 378–395. https://doi.org/10.1111/nyas.13654
- Pearce, M. T., & Wiggins, G. A. (2004). Improved Methods for Statistical Modelling of Monophonic Music. Journal of New Music Research, 33(4), 367–385. https://doi.org/10.1080/0929821052000343840
- Rahali, A., & Akhloufi, M. A. (2023). End-to-End Transformer-Based Models in Textual-Based NLP. AI, 4(1), 54–110. https://doi.org/10.3390/ai4010004
- Rossing, T. D., & Stumpf, F. B. (1998). The Science of Sound. American Journal of Physics. https://doi.org/10.1119/1.12962
- Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., & McIntosh, A. R. (2015). Predictions and the brain: how musical sounds become rewarding. Trends in Cognitive Sciences, 19(2), 86–91. https://doi.org/10.1016/j.tics.2014.12.001
- Sauvé, S. A., & Pearce, M. T. (2019). Information-theoretic Modeling of Perceived Musical Complexity. Music Perception, 37(2), 165–178. https://doi.org/10.1525/mp.2019.37.2.165

- Sloboda, J. A. (1991). Music Structure and Emotional Response: Some Empirical Findings. Psychology of Music, 19(2), 110–120. https://doi.org/10.1177/0305735691192002
- Tay, Y. (2020, 2 mei). Synthesizer: Rethinking Self-Attention in Transformer Models. arXiv.org. https://arxiv.org/abs/2005.00743
- Vaswani, A. (2017). Attention is All you Need. https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
- Zhang, A. (2021, 21 juni). Dive into Deep Learning. arXiv.org. https://arxiv.org/abs/2106.11342

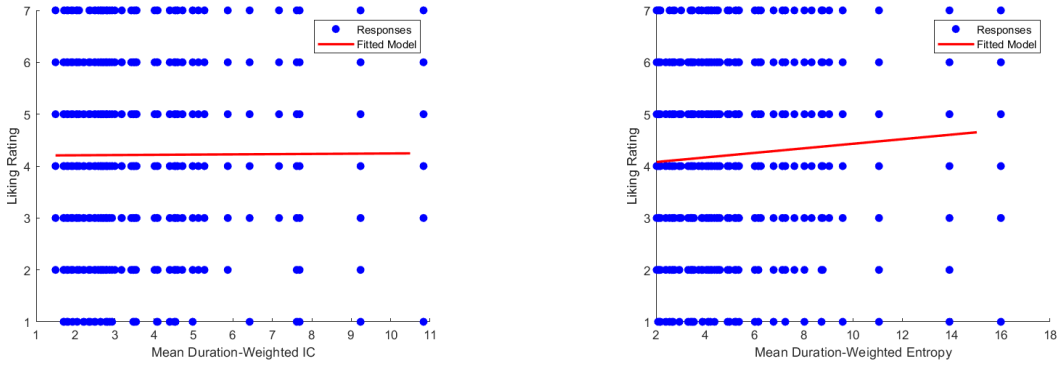
Appendix A

Appendix (optional)



(a) (b)

Figure A.1: **(a)** The optimal model from the IDyOM of mDW-IC by Gold et al. (2019). **(b)** The optimal model from the IDyOM of mDW-Ent by Gold et al. (2019).
The red curve indicates the fitted quadratic model and the blue dots represent the liking ratings from each participant for each stimulus (1906 trials in total).



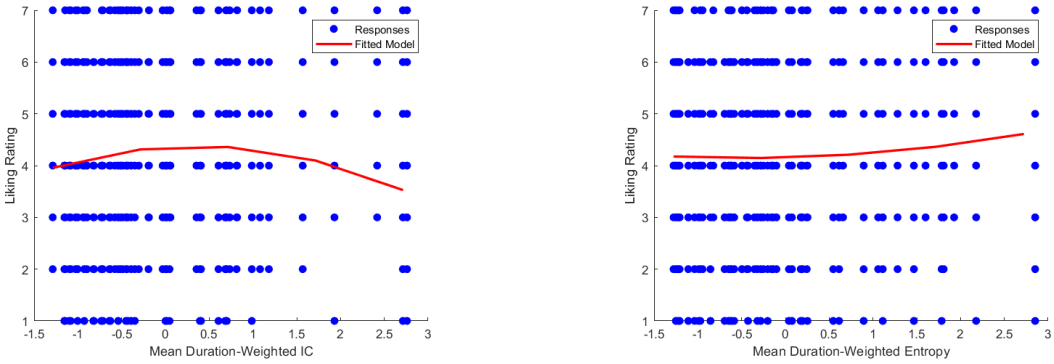
(a) (b)

Figure A.2: **(a)** The model from the MT of mDW-IC *without* the two outliers removed. **(b)** The model from the MT of mDW-Ent *without* the two outliers removed.
The red curve indicates the fitted quadratic model and the blue dots represent the liking ratings from each participant for each stimulus (1906 trials in total).

Piece	Excerpt time (approximate)	Composer	Year	Key	Meter
Streams of Kilnaspig	0:00 – 0:30	Irish traditional	Unknown	G major	Compound duple
Eighteen Studies for the Flute, Op. 41, No. 11	1:30 –2:00	Joachim Andersen	1891	F major	Simple duple
When This Cruel War is Over	1:00 –1:30	American traditional	1863	Bb major	Simple duple
Seven Variations on a Theme from Silvana, J. 128, Op. 33, Var. 7	8:00 – 8:30	Carl Maria von Weber	1854	Bb major	Compound duple
12 Fantasias for Solo Flute, No. 3, Vivace	0:45–1:15	Georg Philipp Telemann	1733	B minor	Simple duple
Eighteen Studies for the Flute, Op. 41, No. 18	0:50 –1:20	Joachim Andersen	1891	F minor	Compound duple
12 Fantasias for Solo Flute, No. 3, Vivace	0:10 – 0:40	Georg Philipp Telemann	1733	B minor	Simple duple
Young Cowherd	0:00 – 0:30	Chinese traditional	Unknown	G major	Simple duple
Sakura	0:00 – 0:30	Japanese traditional	Unknown	D minor	Simple duple
Orchestral Suite No. 2 in B minor, BWV 1067	2:45–3:15	Johann Sebastian Bach	1739	B minor	Simple duple
Eighteen Studies for the Flute, Op. 41, No. 1	0:45–1:15	Joachim Andersen	1891	C major	Simple duple
Five Divertimentos, K. 439b, No. 2, mvmt. 4	0:50 –1:20	Wolfgang Amadeus Mozart	1785	C major	Simple triple
Gavotte	0:00 – 0:30	François-Joseph Gossec	Unknown	C major	Simple duple
Maiden Voyage	2:50 –3:20	Herbie Hancock	1965	F minor	Compound duple
Seven Variations on a Theme from Silvana, J. 128, Op. 33, Theme	0:00 – 0:30	Carl Maria von Weber	1854	Bb major	Compound duple
Drei Fantasiestücke, Op. 73, No. 1	0:30 – 1:00	Robert Schumann	1849	A minor	Simple duple
Five Divertimentos, K. 439b, No. 2, mvmt. 4	3:50 – 4:20	Wolfgang Amadeus Mozart	1785	G major	Simple triple
35 Exercises for Flute, Op. 33, No. 3	1:00 – 1:30	Ernesto Koehler	1880s	F major	Simple triple
Eighteen Studies for the Flute, Op. 41, No. 6	1:00 – 1:30	Joachim Andersen	1891	B minor	Simple triple
Carmen Suite No. 1, Aragonaise	0:45 – 1:15	Georges Bizet	1882	D minor	Simple triple
Orchestral Suite No. 2 in B minor, BWV 1067	0:00 – 0:30	Johann Sebastian Bach	1739	B minor	Simple duple
35 Exercises for Flute, Op. 33, No. 15	0:00 – 0:30	Ernesto Koehler	1880s	E major	Simple duple
Drei Fantasiestücke, Op. 73, No. 1	1:15 – 1:45	Robert Schumann	1849	A minor	Simple duple
Eighteen Studies for the Flute, Op. 41, No. 10	0:00 – 0:30	Joachim Andersen	1891	C# minor	Compound duple
35 Exercises for Flute, Op. 33, No. 10	0:00 – 0:30	Ernesto Koehler	1880s	D major	Simple duple
Study No. 1 in C major, Op. 131	0:00 – 0:30	Giuseppe Gariboldi	1900	C major	Simple duple
Flute Concerto No. 2 in G minor, RV439 "La notte"	10:00 – 10:30	Antonio Vivaldi	1729	C minor	Simple duple

Piece	Excerpt time (approximate)	Composer	Year	Key	Meter
Dolly Suite Op. 56, No. 1	0:10 – 0:40	Gabriel Fauré	1893	G major	Simple duple
Flute Concerto No. 2 in G minor, RV439 "La notte"	9:15 – 9:45	Antonio Vivaldi	1729	G minor	Simple duple
Solo de Concours	4:00 – 4:30	André Messager	1899	Bb major	Simple duple
Student Instrumental Course: Flute Student, Level II book: pg. 12 exercise no. 2	0:10 – 0:40	Douglas Steensland, Fred Weber	2000	Ab major	Simple duple
Eighteen Studies for the Flute, Op. 41, No. 6	0:00 – 0:30	Joachim Andersen	1891	B minor	Simple triple
Fantaisie, Op. 79	0:30 – 1:00	Gabriel Fauré	1898	E minor	Simple triple
12 Fantasias for Solo Flute, No. 5, Allegro	0:37 – 1:17	Georg Philipp Telemann	1733	C major	Simple triple
12 Fantasias for Solo Flute, No. 10, Dolce	1:57 – 2:27	Georg Philipp Telemann	1733	G minor	Simple duple
35 Exercises for Flute, Op. 33, No. 2	0:07 – 0:37	Ernesto Koehler	1880s	G major	Simple duple
12 Fantasias for Solo Flute, No. 10, Presto	2:45 – 3:15	Georg Philipp Telemann	1733	F# minor	Simple triple
Eighteen Studies for the Flute, Op. 41, No. 8	1:30 – 2:00	Joachim Andersen	1891	F# minor	Simple triple
Con Alma	1:15 – 1:45	Dizzy Gillespie	1954	Ab major	Simple duple
35 Exercises for Flute, Op. 33, No. 11	1:00 – 1:30	Ernesto Koehler	1880s	A minor	Compound duple
Syrinx	2:15 – 2:45	Claude Debussy	1913	Bb minor	Simple triple
Orchestral Suite No. 2 in B minor, BWV 1067	3:45 – 4:15	Johann Sebastian Bach	1739	E minor	Simple duple
Nocturnes, Op. 37, No. 1	0:30 – 1:00	Frédéric Chopin	1839	C minor	Simple duple
Seven Early Songs, Die Nachtigall	0:30 – 1:00	Alban Berg	1907	A major	Simple triple
Les Folies d'Espagne, Nos. 7 and 8	0:10 – 0:40	Marin Marais	1701	E minor	Simple triple
Nocturnes, Op. 37, No. 1	0:00 – 0:30	Frédéric Chopin	1839	C minor	Simple duple
Les Folies d'Espagne, No. 5	0:00 – 0:30	Marin Marais	1701	E minor	Simple triple
Le Rossignol en Amour	1:45 – 2:15	François Couperin	1722	G major	Simple triple
Caravan	0:00 – 0:30	Duke Ellington, Juan Tizol	1936	C minor	Simple duple
Citygate/Rumble	1:00 – 1:30	Chick Corea	1986	Db major	Simple duple
First Rhapsody	0:30 – 1:00	Claude Debussy	1910	F# minor, E minor	Simple duple
Alone Together	0:45 – 1:15	Arthur Schwartz	1932	D minor	Simple duple
Seven Early Songs, Traumgekrönt	0:30 – 1:00	Alban Berg	1908	G minor	Simple duple
Les Folies d'Espagne, No. 1	0:00 – 0:30	Marin Marais	1701	E minor	Compound triple
Le Jamf	0:45 – 1:15	Bobby Jaspar	1960	Eb major	Simple duple
Syrinx	0:00 – 0:30	Claude Debussy	1913	Bb minor	Simple triple
Mei	0:37 – 1:07	Kazuo Fukushima	1962	Atonal	Simple duple

Table A.1: Stimulus details for all 57 experimental stimuli from Gold et al. (2019)



(a) (b)

Figure A.3: **(a)** The model from the MT of mDW-IC *with* the two outliers removed. **(b)** The model from the MT of mDW-Ent *with* the two outliers removed.
The red curve indicates the fitted quadratic model and the blue dots represent the liking ratings from each participant for each stimulus (1906 trials in total).