

Seeing Sounds, Hearing Shapes: a gamified study to evaluate sound-sketches

Sebastian Löbbers
Centre for Digital Music
Queen Mary University of London
s.lobbers@qmul.ac.uk

György Fazekas
Centre for Digital Music
Queen Mary University of London
g.fazekas@qmul.ac.uk

ABSTRACT

Sound-shape associations, a subset of cross-modal associations between the auditory and visual domain, have been studied mainly in the context of matching a set of purposefully crafted shapes to sounds. Recent studies have explored how humans represent sound through free-form sketching and how a graphical sketch input could be used for sound production. In this paper, the potential of communicating sound characteristics through these free-form sketches is investigated in a gamified study that was conducted with eighty-two participants at two online exhibition events. The results show that participants managed to recognise sounds at a higher rate than the random baseline would suggest, however it appeared difficult to visually encode nuanced timbral differences.

1. INTRODUCTION

Sample libraries and sound synthesisers are ubiquitous in modern digital music production. However, with samples often organised in lists, finding the “right” sound can be a tedious, frustrating task and, similarly, complex synthesiser controls make it difficult to realise sound ideas in a straightforward way. Tagging of sound material and synthesiser pre-sets improves the search process, but can be imprecise because of the ambiguity of language used to describe sound, or more specifically musical timbre [1, 2, 3]. Sound visualisations can be an additional aid for communicating a sample’s characteristics to a user as seen with waveform representations in digital audio workstations (DAWs) or spectrograms in specialised audio software like Izotope’s *RX* and *Iris*. Further, cross-modal associations between the visual and auditory domain, for example between colours or shapes and sound, can be harnessed for query or perceptually informed visualisations. This paper presents a study that investigates to what extent sound characteristics can be communicated through graphical sketches of personal cross-modal associations. Participants were presented with a gamified survey in which they had to match sounds with sketches that were created in a different study by participants who will be referred to as artists in this paper. The work was conducted in the context of a larger research programme that investigates how

sketches can be used to drive sound synthesis [4]. This section covers relevant research into sound-shape associations and sound visualisations and briefly introduces gamified study design principles. The rest of the paper is organised as follows: Sections 2, 3 and 4 describe the design, analysis and results of the study, a discussion and conclusion are provided in Sections 5 and 6.

1.1 Background on sound-shape associations and sound visualisation

Research suggests that most humans associate sound with elements of the visual domain to some extent [5]. This paper focuses on cross-modal associations between sounds and shapes that were first researched by Wolfgang Köhler who found that the made-up words *takete* and *maluma*¹ are associated with jagged and round shapes. The effect was observed across cultures [6, 7, 8], age groups including toddlers [9], for musical sounds [10] and in a recent study where participants were asked to represent musical sounds through sketching [11]. Sound-shape associations can inform visualisations that communicate sound characteristics without audio playback for example to improve the organisation of one’s personal music library [12], sample selection for live DJ performances [13], exploration of different natural sounds [14, 15] and retrieval of abstract sounds [16]. The idea to use sketches as an input for sound production was first explored with a non-functioning prototype [17]. Further development [18] was made possible through current advancements in deep learning for sketch recognition that is informed by the release of the large-scale sketch dataset *Quick, Draw!* [19].² This study investigates to what extent human participants can recognise sounds from simple sketches to identify perceptually informed features that can help build a deep learning pipeline for sketch-based sound production tasks.

1.2 Study gamification

This study was designed with gamification principles in mind. Research suggests that gamification can have a positive effect on quality and quantity of responses [20] by incorporating the motivational elements of games [21]. For example, participation can be incentivised by a point system [22] with the option to compare results with other “players”³ or by an interface design that adopts a game-like aesthetic as it can be seen in *The Clapping Game* [23] and *Microjam* [24]. The effectiveness of the design can

Copyright: ©2022 Sebastian Löbbers et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

¹ Other studies use *bouba* and *kiki*.

² quickdraw.withgoogle.com/data

³ See Harvard’s *Musical IQ* survey for an example.

be evaluated through user experience survey questions as proposed by Morschheuser et al. [25].

2. METHODS AND MATERIAL

This section describes the design of the study. In ten rounds, one sound and four different sketches were presented of which only one sketch was created with that sound in mind. Participants were tasked to find that match. Each round featured a different sound and sketches from a different artist. The following hypotheses are investigated:

- On average, participants will find matches at a higher rate than the random baseline.
- Participants are most likely to select the sketch corresponding to the next most similar sound when not selecting the correct match.
- The type of sound and an artist's representational sketch style will have an influence on the match rate. The match rate is defined as the frequency at which the correct sound-sketch match is selected.

2.1 Participants

Eighty-two participants took part in the study of which thirty completed an optional post-study survey. The study was part of two online exhibitions that were presented at the Ars Electronica Festival 2020⁴ and the Edinburgh Science Festival 2021.⁵ Sixteen participants completed the study on a mobile device (9 iPhone, 6 Android, 1 iPad) and sixty-six used a desktop or laptop computer (32 Mac, 25 Windows, 5 Linux, 4 unknown). Of the participants who completed the survey 13 were female, 16 male and 1 did not disclose this information. The mean age was 33.6 years ($\sigma = 11.3$). Fourteen stated to actively engage in musical activity at least multiple times a month and two reported a visual impairment (1 short-sighted, 1 near-sighted) and one an auditory impairment (tinnitus).

2.2 Stimuli

All ten sound stimuli and sketches from eleven different artists were selected from a previous study [11] where participants were asked to sketch their personal associations with the sound stimuli. An artist was selected if none of their sketches contained any direct reference to the sound such as depictions of instruments or other recognisable sound sources or symbolic representations like letters or icons. All sketches were created with a MacBook touchpad on a monochromatic digital sketch interface with fixed stroke width. Sounds are equally pitched, loudness normalised and range from musical instruments (*Piano*, *Strings*, *Electric Guitar*) and environmental sounds (*Impact*) to synthesised pads (*Telephonic*, *Subbass*) and abstract textures (*Noise*, *String Grains*, *Crackles*, *Processed Guitar*).⁶

⁴ *The Garden of Forking Paths* at the Ars Electronica Festival 2020

⁵ *Seeing Music* at the Edinburgh Science Festival 2021

⁶ All sounds can be accessed online together with the sketches drawn by participants during the experiment <https://bit.ly/3ta6crU>.

2.3 Apparatus

The study was implemented as a web app in *p5.js* with mobile and desktop capability. The interface was designed with a gamified aesthetic in mind following examples introduced in Section 1.2. The study consisted of ten rounds. In each round, participants were presented with one sound and four sketches of which one was created with that sound in mind and the remaining three in reference to the 3rd, 6th and 9th most dissimilar sound as described in Section 3.1. Sketches were re-drawn in real-time during audio playback following their normalised timestamp data. Participants were asked to find the correct match and after selection, the interface revealed whether their choice was correct. The order of the sounds and artists were randomised on each run. Cookies were used to track returning participants and prevent multiple submissions from the same device.

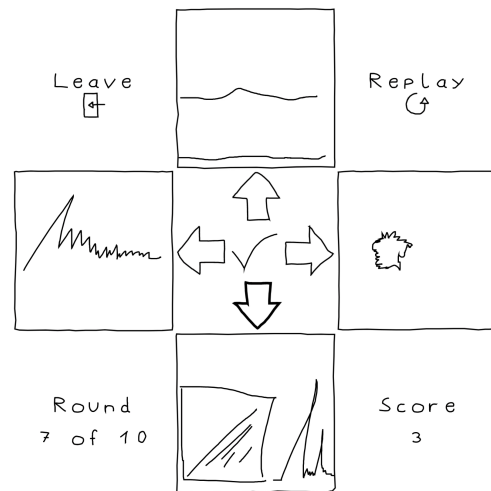


Figure 1. Screenshot of the study interface. The study can be accessed online at <https://phd-studies-eddd5.web.app/>.

2.4 Procedure

In order to make the study more accessible, participants were presented with only one introduction page that included a short description and an animated image to explain how the game works before starting the matching task. Only after completing all rounds, participants were asked for consent to save their data. To incentivise submission, participants could compare their score to other participants after their data had been saved using a mechanism similar to the *Musical IQ* survey described in Section 1.2. Additionally, participants were invited to take part in an optional survey that collected basic personal information (age, gender, visual or auditory impairment), music proficiency using a subset of the GoldMSI survey framework [26] and their experience with the study interface. Information about the device used to complete the study were collected automatically.

3. ANALYSIS

This section provides an overview of the quantitative and qualitative methods used to analyse the study data.

3.1 Sound stimuli dissimilarity

One hypothesis introduced in Section 2 states that participants are most likely to select the sketch corresponding to the next most similar sound when not selecting the correct match. To investigate this, the four sketches in each round relate to the played sound and the 3rd, 6th and 9th most similar sound. To quantify dissimilarity, the euclidean distance between standardised audio feature vectors was calculated similar to [16]. Audio features were extracted with the Librosa library and timbre models [27] as described in [11].⁷

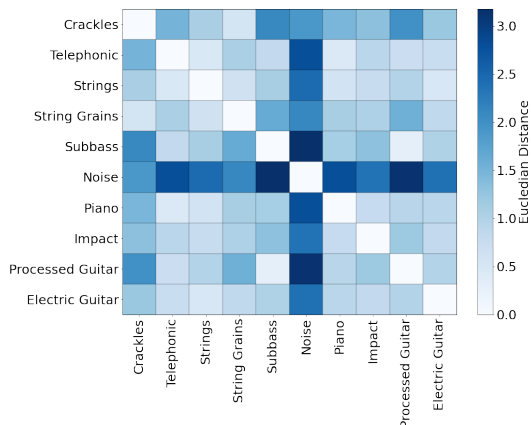


Figure 2. Dissimilarity matrix between sound stimuli. A darker shade indicates a larger euclidean distance between standardised audio features.

3.2 Survey response analysis

In addition to general personal information, the survey contains questions asking whether participants enjoyed the study design. Taking an excerpt from Harms et al.’s survey questions [22], responses were collected using a five point Likert scale on whether they enjoyed playing the game and whether they found the task difficult. In addition, participants were given the opportunity to report any issues that they might have encountered and add any general comments. Likert scale responses were summarised and free-form responses were analysed thematically [28].

3.3 Statistical analyses

The first hypothesis introduced in Section 2 states that on average participants will select correct matches at a higher rate than the random baseline of 2.5 out of 10. Pearson’s Chi-squared test was used to test for significant differences between the collected scores and the random distribution.

Another hypothesis postulates that participants will be more likely to select the sketch of the next most similar sound than the one of the most dissimilar. For this, the match rate for each sound was extracted from the data. However, since the artist and sound order was randomised independently of each other, artists are not evenly distributed among all sounds. Calculating a sound’s match rate by dividing the total number of matches by the total number of occurrences might introduce biases because of an artist’s over-representation. This was mitigated by calculating the match rate for each individual sound-sketch

⁷ A detailed summary of audio feature extraction can be found at <http://doi.org/10.5281/zenodo.4764351>

and ranking them from 10 (highest rate) to 1 (lowest rate) within each artist to obtain the mean rank for each sound. The same procedure was used to compare match rates between artists.

Spearman’s rank correlation was used to investigate whether the mean ranks for sounds correlate with their mean dissimilarity quantified by the mean euclidean distance described in Section 3.1 and whether the mean rank for artists correlates with their sketching style. The sketching style of an artist was quantified by counting the frequency of each of the five categories *Grains*, *Lines*, *Object/Scenes*, *Chaotic/Jagged*, *Radiating* described in [11], however because of the artist filtering described in Section 2.2 the *Object/Scenes* category was removed almost entirely from the subset.

4. RESULTS

The outcome of statistical analysis of the matching task and results of the post-study survey are reported in this section.

4.1 Overall match rate

Figure 3 shows that participants mostly found three (22 participants), four (21 participants) or five (13 participants) correct matches. Chi-square test returned a significant difference between this distribution and the random baseline described in Section 3.3 ($\chi^2(1, N = 82) = 99509.11$ $p < .0001$). In total, the correct option was selected most frequently (307/820 selections), however the second most frequent selection was the option that refers to the most dissimilar sound (212/820). These results suggest that participants were able to extract information from the sketches that helped them to find correct matches, however, contrary to expectation, participants did not predominately choose the sketch referring to the next most similar sound when not finding the correct match.

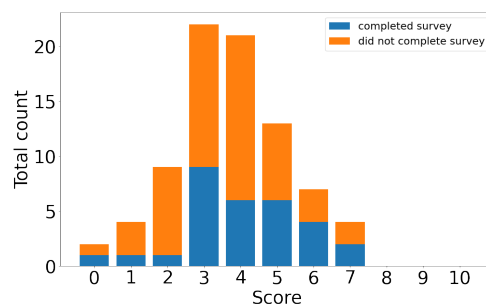


Figure 3. Match rate of all participants. The mean match rate is 3.74 for all participants, 4.07 for participants who completed the survey and 3.56 for participants who did not complete the survey.

4.2 Match rate by sound

Table 1 shows how often each of the four sketch options were selected for each sound. Except for *Telephonic* (18/82 selections) and *String Grains* (22/82) the correct option was selected most frequently with *Noise* (49/82) and *Piano* (47/82) showing the highest match rate. Figure 2 shows that with an average euclidean distance of 2.57 *Noise* is the most dissimilar sound to all other sounds except for

Crackles which is the second most dissimilar sound with a euclidean distance of 1.46. These sounds also show a high mean match rank 7.45 for *Noise* ($\sigma = 2.96$) and 6.4 for *Crackles* ($\sigma = 2.27$) compared to sounds that are less dissimilar like *Telephonic* with a euclidean distance of 1.06 and a mean rank of 3.23 ($\sigma = 1.95$) or *Strings* with a euclidean distance of 0.96 and a mean rank of 4.27 ($\sigma = 2.89$). Spearman rank correlation suggests a strong positive correlation between the mean dissimilarity of a sound and its mean score rank ($\rho = .7$, $p < .05$) as visualised with a trend line in Figure 4. However, the figure also shows that *Piano* has the second highest mean rank ($\mu = 7.32$, $\sigma = 2.25$) despite a mid-range dissimilarity (euclidean distance of 1.13). These results suggest that some sketches are more likely to be correctly matched to a sound which is at least partially influenced by the sound's mean dissimilarity to the remaining stimuli.

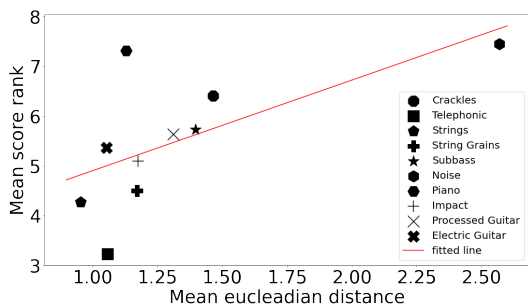


Figure 4. Correlation between match rate rank by sound and sound dissimilarity quantised by the mean euclidean distance between audio features.

4.3 Match rate by artist

The sketching style of an artist appears to have an effect on how well their sketches are matched by participants. Using the sketch categories described in Section 3.3 a significant correlation could be found between the frequency of sketches in the *Radiating* category and the mean rank for each artist ($\rho = -.71$, $p < .05$). This trend can also be observed when comparing categories between the sketches with the highest and lowest match rate for each sound seen in Table 2 with *Radiating* occurring 7/10 times for the worst performing sketches, but only 3/10 times for the best performing ones. It has to be noted that from all categories *Radiating* is spread the most evenly across sounds (Cramér's V: .16).

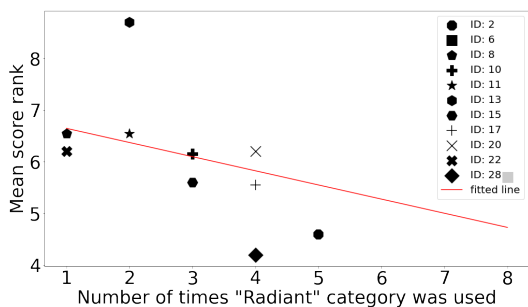


Figure 5. Correlation between match rate rank by artist and prevalence of *Radiating* category in sketches by an artist.

4.4 Study Design Evaluation

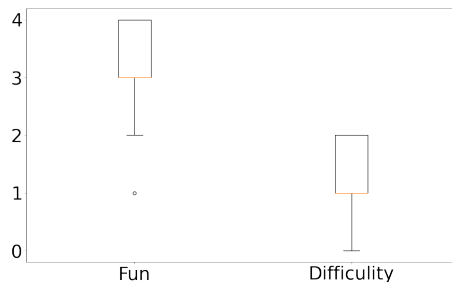


Figure 6. Answers to the questions *I had fun playing the game.* (0-completely disagree to 4-completely agree) and *How difficult was the game?* (0-very difficult to 4-very easy)

A secondary question of this study was to investigate the effectiveness of the gamified study design. Eighty-two participants were recruited from visitors of the online exhibitions. The age range of participants described in Section 2.1 indicates that a diverse demographic could be reached. Figure 6 shows that despite finding the task difficult most participants enjoyed completing the study. The responses for *having fun* align with findings by [22]. The free-form answers reveal that the score system of the study design was criticised by some participants finding it “difficult getting a ‘score’ for a subjective agreement on sound” (Participant X1) or not being “sure how you define the ‘correct’ answer” (Participants X1). Another point of criticism was the short introduction section with participants saying that “the instructions went by too quickly” (Participant X3), the “interface was tricky to understand” (Participant X4) and one suggested to “have like a test sound before the game starts” (Participant X5).

5. DISCUSSION

The gamified study design and integration into non-academic exhibitions proved successful at reaching a large and varied participant pool without high recruitment effort. However, spanning over the course of a year, data collection was slow. According to survey responses analysed in Section 4.4, participants enjoyed completing the study even though they perceived the task to be difficult. The score system of the design appeared to be confusing to some with survey responses indicating that scores might not be adequate for this task. Some participants seemed to struggle with understanding the interface and the task possibly because the introduction was short and easy to skip. A tutorial or training run could be included to make sure that all participants understand the task. While the integration into an online exhibition might lower initial resistance to participate, it can make it more difficult to motivate participants to complete more tedious tasks like filling in a questionnaire. Figure 3 shows that a majority of participants did not continue with the post-study survey after submitting their score. Overall, a gamified design presented in a non-academic context might be most suitable for studies that are not time-sensitive and do not require a tightly controlled study environment. Results presented in Section 4.1 show that participants found the

Correct Option	2. Option (3 rd distance)	3. Option (6 th distance)	4. Option (9 th distance)
Crackles: 30	Electric Guitar: 10	Telephonic: 24	Subbass: 18
Telephonic: 18	Processed Guitar: 22	Impact: 18	Noise: 24
Strings: 28	Piano: 14	Processed Guitar: 24	Noise: 16
String Grains: 22	Electric Guitar: 13	Piano: 4	Noise: 43
Subbass: 27	Electric Guitar: 15	Impact: 21	Noise: 19
Noise: 49	Impact: 13	Telephonic: 4	Subbass: 16
Piano: 47	Impact: 17	String Grains: 7	Noise: 11
Impact: 32	Electric Guitar: 19	Processed Guitar: 11	Noise: 20
Processed Guitar: 29	Piano: 7	Impact: 24	Noise: 22
Electric Guitar: 25	Impact: 23	Processed Guitar: 11	Noise: 23
Total: 307	Total: 153	Total: 148	Total: 212

Table 1. Match rates for each sound. For most sounds the correct option was selected most of the times. However, some sounds were often mismatched (e.g. String Grains and Noise).





















Sound Name	Crackles	Telephonic	Strings	String Grains	Subbass	Noise	Piano	Impact	Proc. Guitar	Electric Guitar
Best	 6/11	 <u>7/12</u>	 9/15	 <u>4/9</u>	 4/9	 10/12	 8/10	 9/12	 6/12	 <u>4/9</u>
Worst	 <u>1/9</u>	 <u>1/9</u>	 <u>3/10</u>	 0/9	 <u>0/13</u>	 <u>5/9</u>	 6/10	 2/9	 <u>2/10</u>	 <u>1/10</u>

Table 2. Sketches with the best and worst match rates. To reduce the risk of skewing the data through small sample sizes, only sketches were considered that were shown at least 8 times which is the median number of occurrences. Underlined sketches are part of the *Radiating* category.

correct sound-sketch matches at significantly higher rate than the random baseline. However, the match rate is notably lower than the results of similar studies that use visual stimuli that were purposefully crafted by a single artist [16, 13]. This is not unexpected as the sketches used in this study were not produced with the goal to maximise recognition for other people. Instead the results suggest that even sketches that were produced intuitively in a short time can communicate information to identify the sound that they represent. The results further show that match rates differ among different sounds and artists. Sketches from sounds that are more dissimilar to the remaining stimuli are more frequently matched correctly. This suggests that sound-sketches can successfully communicate general sound information but fail to encode more detailed differences in timbre. Against expectation, participants were most likely to select the sketch relating to the most dissimilar sound rather than the next most similar one when not finding the correct match. This could mean that participants adopted a selection process that identifies the “odd one out” when three of the four sketches looked similar. Analysis of sketching category frequencies presented in Section 4.3 indicate that some sketching styles might be better at conveying information about sound than others, but as noted this could be caused by a more even distribution of some styles across sounds which makes them more likely to be mismatched than a style that is more strongly associated with a specific sound. In addition, outliers to these trends exist for sounds and artists potentially because the quantified measures for sound dissimilarity and sketching style cannot capture all dimensions of human percep-

tion. Qualitatively reviewing the results suggests that the perceptual dimension from noisiness/complexity to calmness/simplicity might be encoded especially well in the sketch representations which would explain the high mean match rank for *Noise* and *Piano*.

6. CONCLUSION

This study shows that graphical sketch representations of sounds can be recognised successfully by human participants if their timbres are sufficiently different from each other, but it appears to be difficult to visually encode timbral differences of similar sounds. In the context of developing a sketch-based sound synthesiser, a deep learning mapping model might be limited to predicting general sound categories even when trained on a larger sound-sketch dataset. Similar studies that reported higher matching rates used purposefully crafted visual stimuli suggesting that introducing guidelines on how to represent different perceptual dimensions could increase a model’s accuracy. A different approach could be found in tuning a base model according to an individual user’s sketching style.

Acknowledgments

EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology (EP/L01632X/1).

7. REFERENCES

- [1] C. Saitis, S. Weinzierl, K. von Kriegstein, S. Ystad, and C. Cuskley, “Timbre semantics through the lens

- of crossmodal correspondences: A new way of asking old questions,” *Acoustical Science and Technology*, vol. 41, no. 1, pp. 365–368, 2020.
- [2] K. Siedenburg, I. Fujinaga, and S. McAdams, “A comparison of approaches to timbre descriptors in music information retrieval and music psychology,” *Journal of New Music Research*, vol. 45, no. 1, pp. 27–41, 2016.
- [3] S. McAdams and B. L. Giordano, “The perception of musical timbre,” *The Oxford handbook of music psychology*, pp. 72–80, 2009.
- [4] S. Löbbers and G. Fazekas, “Sketching Sounds: Using Sound-shape Associations to Build a Sketch-based Sound Synthesizer,” in *Digital Music Research Network (DMRN+16)*, 2021.
- [5] G. Martino and L. E. Marks, “Synesthesia: Strong and weak,” *Current Directions in Psychological Science*, vol. 10, no. 2, pp. 61–65, 2001.
- [6] R. Davis, “The Fitness of Names to Drawings. a Cross-Cultural Study in Tanganyika,” *British Journal of Psychology*, vol. 52, no. 3, pp. 259–268, 1961.
- [7] I. K. Taylor and M. M. Taylor, “Phonetic Symbolism in Four Unrelated Languages,” *Canadian Journal of Psychology/Revue canadienne de psychologie*, vol. 16, no. 4, pp. 344–356, 1962.
- [8] A. J. Bremner, S. Caparos, J. Davidoff, J. de Fockert, K. J. Linnell, and C. Spence, ““Bouba” and “Kiki” in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches to Westerners,” *Cognition*, vol. 126, no. 2, pp. 165–172, 2013.
- [9] D. Maurer, T. Pathman, and C. J. Mondloch, “The Shape of Boubas: Sound–Shape Correspondences in Toddlers and Adults,” *Developmental Science*, vol. 9, no. 3, pp. 316–322, 2006.
- [10] M. Adeli, J. Rouat, and S. Molotchnikoff, “Audiovisual correspondence between musical timbre and visual shapes,” *Frontiers in human neuroscience*, vol. 8, p. 352, 2014.
- [11] S. Löbbers, M. Barthet, and G. Fazekas, “Sketching sounds: an exploratory study on sound-shape associations,” in *International Computer Music Conference (ICMC)*, 2021, p. 6.
- [12] P. Kolhoff, J. Preub, and J. Loviscach, “Music Icons: Procedural Glyphs for Audio Files,” in *2006 19th Brazilian Symposium on Computer Graphics and Image Processing*, Oct. 2006, pp. 289–296.
- [13] Y.-X. Chen and R. Klüber, “ThumbnailDJ: Visual Thumbnails of Music Content.” in *11th International Society for Music Information Retrieval Conference (ISMIR)*, 2010, pp. 565–570.
- [14] C.-H. Wan, S.-P. Chuang, and H.-Y. Lee, “Towards audio to scene image synthesis using generative adversarial network,” in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 496–500.
- [15] T. Ishibashi, Y. Nakao, and Y. Sugano, “Investigating audio data visualization for interactive sound recognition,” in *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 2020, pp. 67–77.
- [16] T. Grill and A. Flexer, “Visualization of Perceptual Qualities in Textural Sounds,” in *International Computer Music Conference (ICMC)*, 2012, p. 8.
- [17] P. Knees and K. Andersen, “Searching for audio by sketching mental images of sound: A brave new idea for audio retrieval in creative music production,” in *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, 2016, pp. 95–102.
- [18] L. Engeln, N. L. Le, M. McGinity, and R. Groh, “Similarity Analysis of Visual Sketch-based Search for Sounds,” in *Audio Mostly 2021*, 2021, pp. 101–108.
- [19] D. Ha and D. Eck, “A Neural Representation of Sketch Drawings,” *arXiv:1704.03477 [cs, stat]*, May 2017.
- [20] J. Hamari, J. Koivisto, and H. Sarsa, “Does Gamification Work? – A Literature Review of Empirical Studies on Gamification,” in *2014 47th Hawaii International Conference on System Sciences*, Jan. 2014, pp. 3025–3034.
- [21] A. Rapp, F. Hopfgartner, J. Hamari, C. Linehan, and F. Cena, “Strengthening Gamification Studies: Current Trends and Future Opportunities of Gamification Research,” *International Journal of Human-Computer Studies*, vol. 127, pp. 1–6, Jul. 2019.
- [22] J. Harms, S. Biegler, C. Wimmer, K. Kappel, and T. Grechenig, “Gamification of online surveys: Design process, case study, and evaluation,” in *IFIP Conference on Human-Computer Interaction*. Springer, 2015, pp. 219–236.
- [23] S. Duffy and M. Pearce, “What makes rhythms hard to perform? An investigation using Steve Reich’s Clapping Music,” *Plos one*, vol. 13, no. 10, p. e0205847, 2018.
- [24] C. P. Martin and J. Torresen, “Data-Driven Analysis of Tiny Touchscreen Performance with MicroJam,” *Computer Music Journal*, vol. 43, no. 4, pp. 41–57, 2020.
- [25] B. Morschheuser, J. Hamari, K. Werder, and J. Abe, “How to Gamify? A Method For Designing Gamification,” in *Hawaii International Conference on System Sciences*, 2017.
- [26] D. Müllensiefen, B. Gingras, J. Musil, and L. Stewart, “The musicality of non-musicians: an index for assessing musical sophistication in the general population,” *PloS one*, vol. 9, no. 2, p. e89642, 2014.
- [27] A. Pearce, T. Brookes, and R. Mason, “Modelling Timbral Hardness,” *Applied Sciences*, vol. 9, no. 3, p. 466, 2019.
- [28] V. Braun and V. Clarke, “Using Thematic Analysis in Psychology,” *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006.