‘Mood-Zart’: A Tangible Musical Interface Designed to Help Convey Emotion

MATT DIXON, DYLAN HILL, ANI MOHANAMURALI, YASHVEER RAI BASGEET, and ALFIE YUNNIE

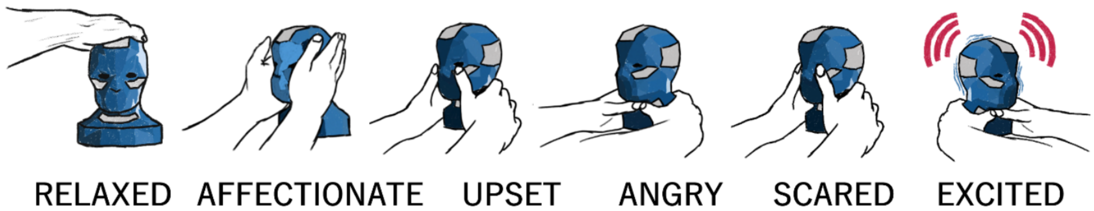

Fig. 1. Emotions recognised by Moodzart, 2023.

Expressing and capturing emotion is fundamental to the process of making music. The technical knowledge required to use traditional audio manipulation tools, such as DAWs (Digital Audio Workstations), can act as a barrier to musicians who lack extensive knowledge of software operations, sound design and music theory. In addition, the time taken to translate ideas into audio output using current tools can restrict the ability to capture the emotions present ‘in the moment’. In this paper, we present ‘Mood-Zart’, a musical interface resembling a human head that affords emotional interaction and provides a novel audio engineering experience. Built primarily for musicians and producers to quickly modify and develop sonic ideas, ‘Mood-Zart’ helps individuals actualise their ideas more efficiently without breaking flow states induced by composition. The human resemblance acts as an abstraction of technicalities and provides an intuitive interface for conveying emotion. Users can interact with touches, squeezes, and various other gestures to provoke a corresponding transformation to their provided audio signal. The device increases accessibility by reducing skill requirements, introducing a new set of individuals to the world of music production whilst streamlining the formation of creative ideas.

CCS Concepts: • **Human-centered computing → Sound-based input / output**; **Haptic devices**.

Additional Key Words and Phrases: emotion recognition, multisensory input, touch detection, audio processing

ACM Reference Format:
Matt Dixon, Dylan Hill, Ani Mohanamurali, Yashveer Rai Basgeet, and Alfie Yunnie. 2023. ‘Mood-Zart’: A Tangible Musical Interface Designed to Help Convey Emotion . In *COMS3055: Interactive Devices.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The increasing availability of music production technology has inevitably led to more individuals recording, editing, and arranging music through software and hardware tools [Wanderley and Orio 2002]. Whilst primary tools for audio manipulation, such as DAWs (Digital Audio Workstations), provide useful abstractions of the technical sonic processes taking place, research has identified that the user-interface metaphors from the past no longer support the current activities of professional producers [Duignan et al. 2010]. Moreover, these digital representations are meaningless to users who lack prior knowledge of the recording studio hardware on which these metaphors are based and can actively

2023. Manuscript submitted to ACM

be a barrier to the identified central goal of producing music: to effectively express emotion through sound [Lopez et al. 2010].

Research into the development of new music interfaces has shifted its attention to usability as a core principle when designing both physical and virtual interfaces. In turn, key HCI (Human Computer Interaction) principles have become a common point of reference to best understand how user-friendly systems with clear affordances can be developed. [Wanderley and Orio 2002] highlight how when designing input devices for musical expression, there is a tendency to either base design on pre-existing models of musical interaction or deliberately avoid any relationship to familiar gestural models. Whilst design based on acoustic instruments or industry standard controllers may aid the initial operational understanding of skilled users, methods of conveying emotion remain ambiguous regardless of a user's prior experience. Recurrently creating models based on the principle of familiarity will further feed the status quo bias of current interface designs - research clearly highlights there are improvements to be made [Duignan et al. 2010]. On the other hand, idiosyncrasy is highlighted by [Jordà 2002] as the biggest problem preventing new musical controllers from being accessible to wider audiences. Musical interfaces must not be so widely abstract that users are left confused about usage as this will likely lead to a lack of adoption in regular practice. Musical tasks should generally strive for maximal simplicity [Wanderley and Orio 2002] since level of engagement and degree of satisfaction are rooted from a clear understanding of the device at hand [Tahiroğlu et al. 2017]. There is, therefore, clear scope for exploration into new models that are able to afford universal emotional interactions to users of varying expertise.

[Duignan et al. 2010] acknowledge pivotal points of breakdown that take place in the current typical workflow of music producers and composers using DAWs. The 'editing performance divide' is the disruption caused when a musician is forced to switch from using a physical MIDI or performance instrument to a virtual interface to continue the mixing and manipulation of audio. Participants from their study indicate that DAWs are not "really made to improvise on" and that when "jamming" the abrupt starting and stopping needed to make changes can be disruptive to the process of making music. The software tools used to modulate audio are stated to be "mentally intensive" due to the tedious task of tweaking fine details. The authors suggest that abstraction of processing should be considered more heavily in the design of future DAWs. In addition, in typical working scenarios, audio modulation takes place offline, with [Micallef Grimaud and Eerola 2020] highlighting that previous attempts at creating emotional music interfaces provide no way to hear the impact of the changes made live. Real-time audio feedback is essential for maintaining flow experiences for musicians [Bloom and Skutnick-Henley 2005] so incapability to provide an instantaneous response will undoubtedly cause hinderance. There is a clear lack of intuitive live audio modulation hardware that can be used as an alternative to the current mediocre practices using DAWs.

Embodied agents are a growing area of research in interface design, as they provide a potential way to enhance human computer interaction [Creed and Beale [n. d.]]. Humanistic features can take advantage of pre-existing social skills and provide natural, engaging interactions. The 'Skin-On' interface designed by [Teyssier [n. d.]] garnered interest due to its unique texture and innovative use of 'skin' to provide novel gestures. The interest surrounding the device highlights the desire for more natural and interesting methods of input. Whilst it should be mentioned that opinions regarding embodied agents are mixed, predominant negative views stem from agents being of poor quality or cheap virtual imitations of human social interaction – research fails to highlight the potential of tactile agents [Creed and Beale [n. d.]]. The role these devices will play in future contexts is uncertain and consensus remains that technical and social issues need to be considered, however since current research primarily focuses on digital agents, more research into tangible anthropomorphised devices could help inform decisions regarding emotion when designing hardware as a general principle.

We propose 'Mood-Zart' as a device to help facilitate the expression of emotion in music. Skilled traditional musicians intuitively know how to use their instruments as tools for expression, allowing them to enter states of flow [Bloom and Skutnick-Henley 2005], however, when it comes to editing and adding effects to a recorded audio signal, they lack an easy tangible method to quickly generate ideas. We aim to provide this as a universal capability to users regardless of their technical knowledge and hence enhance the experience of music production. The 'head' interface is simple enough that it affords interaction whilst also providing the ability to change audio input, adding novelty and encouraging user exploration. We believe that with the right balance of familiarity and complexity, the device will be able to strengthen and sustain flow experiences activated through the process of making music.

2 RELATED WORK

2.1 Capturing Emotion in Real-Time

"EmoteControl" is a software tool developed to help users with no prior skills convey different emotions by changing the expressive and structural cues of a MIDI input [Micallef Grimaud and Eerola 2020]. Users can make real-time changes to tempo, pitch, dynamics, articulation, brightness, and mode with sliders on a graphical user interface. The advantage of allowing users to work in real-time is that they can continuously adjust the sound till they achieve their desired output and, additionally, this serves as an educational tool teaching the user how musical attributes impact the sound generated. Feedback from the device's evaluation critiqued the interface's design and its use of "bleak colours". When asked what participants would change, colour and design were recurring themes, with some participants suggesting the inclusion of term definitions for non-musicians if the intended purpose is education - users were unclear of what the names of musical cues meant. Positive remarks included participants stating the changes they made using the software could be easily heard and many mentioned they felt successful in conveying emotions.

The "Emotion-Driven Music Engine" is a system that aims to produce music expressing specified emotions [Lopez et al. 2010]. It is highly modular with multiple parameters being changed in accordance with the input of an emotional description. Suitable segments of audio are selected from pre-classified existing music to form song-like structures representing the intended emotion. The input can also be adapted depending on the use case. For example, when used as part of an interactive art installation that changed emotion based on the presence of people, the input was movement data from sensors monitoring the number of people in the room [Ventura et al. 2009]. The installation provided real time visual feedback in addition to audio output, which proved useful for users understanding what emotion was trying to be conveyed and allowed for an immersive experience.

[Desai et al. 2018] developed "Mindtrack": a brain-computer interface that uses brainwave data to allow users to expressively shape progressive music. This device requires a complex electroencephalogram headset that monitors user brain activity. Detected emotions were categorized into very sad, sad, neutral, happy, or very happy, but it is implied there is scope for extension. The device is primarily beneficial to people with physical impairments as it allows them to express themselves musically where otherwise may be difficult. The main limitation of a system like this is the lack of user control. When audio is purely generated through a fixed sensor, there is potential for misreading of brainwaves and the audio output may not match the intended emotion of the user. Despite the complex technology underpinning the input, the output generated from this system is minimal.

Despite there being positive feedback and a beneficial use case for each of these systems, there are clear gaps in research that we are aiming to fill with our device. Firstly, all the aforementioned systems make changes to MIDI data or use pre-determined sounds, reducing the flexibility of audio input. [Micallef Grimaud and Eerola 2020] highlight that there are limitations to the changes that can be made in an audio format, as structural cues such as articulation cannot be easily modified outside of MIDI. With most producers converting their MIDI data to audio before the editing stage anyway [Duignan et al. 2010], audio input of the user's selection would create a more universal tool. We designed 'Mood-Zart' to receive and manipulate audio signals rather than MIDI for this very reason. In addition, there are restrictions imposed by the visual appearance of "EmoteControl". Whilst the interface was easy to use, users complained about aesthetics [Micallef Grimaud and Eerola 2020]. 'Mood-Zart' accounts for this by having a visually stimulating interface with an intriguing shape and bright colours to help sustain interest. Since "EmoteControl" functions primarily as an educational tool for non-skilled users rather than focussing on experience, users may waste time by making minute adjustments to sounds instead of quickly getting ideas actualised. 'Mood-Zart' does not require the user to know about features like 'tempo' or 'pitch' but rather the emotion they want to convey. "Mindtrack" and applications of the "EDME" are commendable in this sense as they make use of process abstraction, so users worry less about the technicalities of system output [Lopez et al. 2010]. We decided that abstracting away musical jargon and instead utilizing emotional and visual cues would provide a more accessible experience to users of all skill levels.

Each of these devices successfully use real-time processing to help user's hearing the consequence of their interactions. 'Mood-Zart' follows suit by producing live audio feedback based on the user input into the audio workstation of choice. One thing that our system provides that can be seen to be lacking in those mentioned above is the clarity of intention behind action. Percieved affordances are essential for users to understand the possible actions they can take in a task [Magnusson 2010]. Without them, users are left clueless to potential interactions with a system. The systems above rely on trial and error for users to achieve the sound they intend. In time-precious situations, such as during the conception of creative ideas, there may not always be scope for errors to be made [Duignan et al. 2010]. The head shape of our system makes interaction intuitive, allowing for a smoother process to reach desired output.

2.2 Tangible Music Interfaces

Both [Winkler 1995] and [Camurri et al. 2000] investigated the use of movement and gesture mapping in intelligent interactive music systems. The human body as an instrument is free from any associations that may be present with typical acoustic instruments and therefore there are different characteristics to be acknowledged when designing an interface around it. [Winkler 1995] states that interactions must reflect weight, force, pressure, speed and range when producing a sound, encapsulating the effort and energy used to create it. With 'Mood-Zart', we planned to keep this in mind when considering how best to detect interactions.

3 USER INTERACTION STUDY

Although the device aims to alter audio characteristics to reflect emotion, countless previous studies focussed on correlating emotions to varying expressive and structural cues in music have come to the same conclusions ([Livingstone et al. 2010]; [Ventura et al. 2009]; [Desai et al. 2018]; [Wassermann et al. 2003]). Emotions can be mapped to discrete segments of a valence-arousal graph which is used to determine attributes of music such as tempo, pitch, and articulation. We relied on existing research to inform our decisions regarding audio manipulation. It is worthy of noting that despite humans being capable of expressing multitudes of emotion on varying scales, there is a limited number that can be effectively identified in music with consensus agreement [Juslin and Laukka 2003].

[El-Shimy and Cooperstock 2016] acknowledge that new musical interfaces struggle to be researched using traditional HCI methods, as the "goals", "tasks" and "needs" of users can be more ambiguous with non-utilitarian devices. To mitigate against this, they outline three key principles regarding research of these interfaces; the first being to ensure basic interactions are understood and validated. This is particularly beneficial with interfaces of unconventional form. With the planned atypical shape and novel input format of our ‘head’, our primary goal of this study was to understand the potential interactions a user may intend to impose. The study was performed with fifteen people.

Equipping the participant with a Styrofoam model human head, we verbally stated emotions and instructed that they attempt to inflict this emotion on the head in whichever way they saw fit. The selected test emotions were relaxed, excited, affectionate, angry, upset, scared and nervous – highlighted from the circumplex model of affect [Posner et al. 2005]. A designated observer took qualitative notes of each participant’s interactions that we consequently clustered to collate common themes.

On analysis, we segmented the head to identify where the highest frequency of interactions were present and what action was taken to inflict them. Participants made use of gestures ranging from touching to stroking to pinching to attempts at verbal communication with the device. We were able to identify three main contact points – the top of head, back of head and cheeks. When instructed to make the head feel ‘relaxed’, 66% of participants interacted with one of these listed areas. The most common action was stroking with 66% of participants performing a variation at some point during the study. Shaking of the head was frequently used to represent more activated emotions – 53% of participants made us of this action.

4 DESIGN AND IMPLEMENTATION

4.1 Housing Construction

The shell uses 3D-printed PLA with a Gyroid infill pattern, to maintain stability whilst reducing print time. The 3D model for the head was an altered version of a free-to-use model from Turbosquid, by [janovich [n. d.]]. We altered the geometry around the neck and cheek areas to better fit our silicone pads, and created divots for each of them. The specific placement of the pads identifies the interaction points across the device, and creates an interesting texture for pressing and stroking. The silicone pads were created using separate 3D printed moulds, then placed into the shell.

4.2 Device Hardware

We decided to build our device around a Teensy 3.6 Development Board, as this device had all of the key functionality we required. We used the Teensyduino extension for the Arduino IDE, which supports I2C connectivity for external devices and has multiple pins with dedicated support for capacitive touch sensing.

To create the touch-pads themselves we applied conductive copper tape at specified points on the face, as decided from the user interaction study, and covered them with the silicone pads. For pressure-sensing we used Velostat sheet. It can be used to detect force on its surface because the conductivity changes when the pressure on the material changes, as the resistance lowers.

Other input signals were collected from external sensors. To detect impact and movement of the device we used an Adafruit ADXL343 accelerometer, as it could report these events simultaneously through the use of a standard synchronised I2C connection and an interrupt pin. To detect if the eye of the device is being covered, we used an Adafruit APDS9960 Proximity and Gesture Sensor which we configured to continually output the relative distance of an object to the sensor over the same I2C connection.

Fig. 2. Completed device

4.3 Software

When writing the software for 'Mood-Zart' making sure that it integrated smoothly with the hardware in order to deliver a responsive and cohesive experience was a priority. The idea was that whenever the user would interact with the interface the gesture was immediately identified and assigned to the corresponding emotion. We also wanted to represent those changes on a Circumplex model graph. To map the corresponding emotions to a certain gesture, the software constantly monitors all the sensors around and within the device. This helps ensure that 'Mood-Zart' is constantly checking for any possible change in the sensor readings thus ensuring quick and efficient responses. Here, the software is integrating four types of sensors: a proximity sensor; an accelerometer; touch sensors and pressure sensors. These sensors were created based on the interactions observed in our User Interaction study, to ensure gestures felt intuitive and natural. The proximity sensor was calibrated to detect when users placed their hand in front of the eyes and also accounted for detecting the internal structure of our interface so as to avoid any false positives. Similarly, the touch sensors had to be calibrated in such a way to be responsive once a silicon layer is applied over the copper tape. The accelerometer was dependent on the motion of 'Mood-Zart' as a whole in a repetitive upward and downward motion to simulate a shaking motion. Lastly, the pressure sensor was calibrated in such a way as to incentivise the user to apply a reasonable amount of force to the neck area thus properly conveying the 'anger' emotion. The software allows visualisation of the changes on a Circumplex graph in real-time as the user interacts with 'Mood-Zart'. This is achieved by incrementing or decrementing the arousal and valence value depending on the emotion being detected. The software also ensures that the increase/decrease happened gradually and not instantly when an emotion is detected. This helps the user to be more in control of the level of emotion and subsequent effects they want to apply to their music. The software allows for more than one emotion to be detected at the same time and the graph to be adjusted accordingly to represent both emotions, thus adding another layer of customisation to the process. Initially, when an interaction was detected, the relevant emotion and Circumplex coordinate was displayed in plain text on the device's serial output.

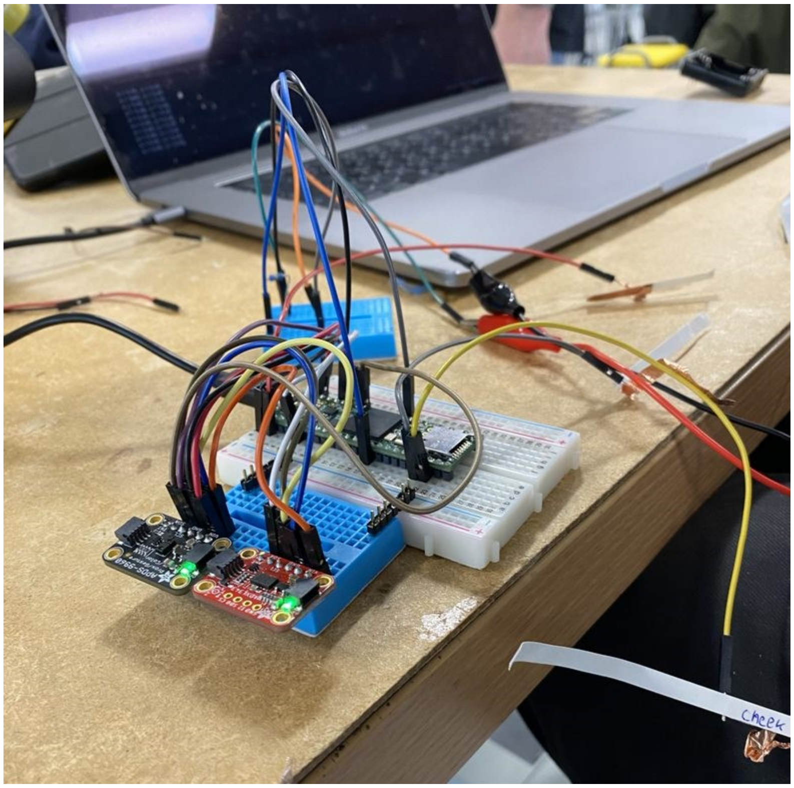

Fig. 3. External sensor testing

However, we found that users experienced difficulty recognising which emotion they had triggered. Therefore, ASCII art was used to allow for a more comprehensible understanding of the device’s evaluation of interactions. In addition, displaying the changes in the coordinates of the Circumplex graph helps to show how the emotion was changing.

4.4 Emotions and Musicality

The emotions we used were based on the Circumplex Model of Affect [Posner et al. 2005]. This allowed us to map emotions and musical effects across a two-dimensional space. Using a continuous representation of emotions allows continuous change of musical effects, enabling more nuanced expression for the user.

Several different features of music and musical effects and the emotions that they elicit were utilised, including loudness, pitch, rhythm, timbre, and tonality [Gabrielsson and Lindström 1993]. We also used effects that could be linked to groups of emotions, based on [Zentner et al. 2008] paper on ‘Emotions Evoked by the Sound of Music’. Louder music is associated with higher activation emotions, whilst quieter music is associated with lower activation. Higher pitches are also associated with higher activation, and lower pitches with lower activation, and additionally with lower valence. Increased modulation is associated with higher activation, and lower valence. Irregular rhythms are associated with high activation in combination with low valence. In general, ‘softer’ sounds are for lower activation, and ‘harsher’ sounds are for higher activation. There is more irregularity towards lower valence, and more structure and form towards

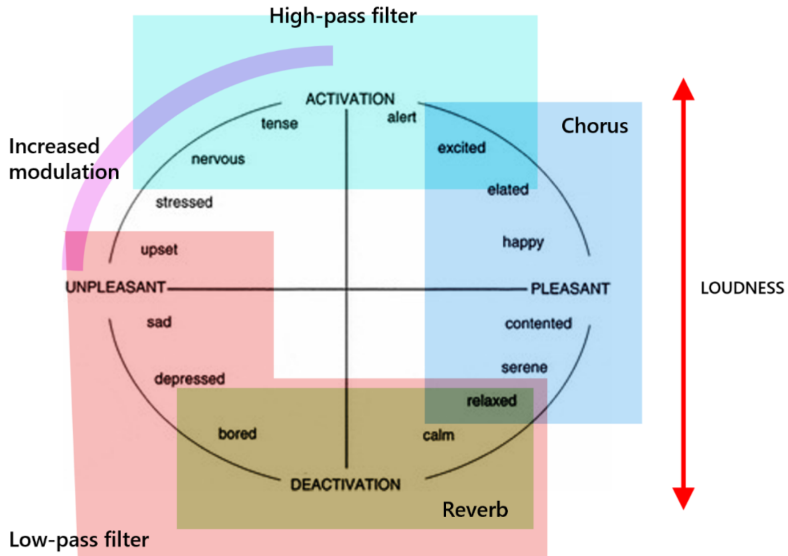

Fig. 4. Audio Effects relating to different sections of the Valence Arousal model

higher valence. This resulted in the creation of a map of the Circumplex Model to dictate how different effects would be applied to music based on the desired emotion's position on the model. Pictured in Figure 4 is a simple version, though this could be made more complex and detailed if desired, however as mentioned this was not the main focus of this project.

5 STUDY 2: DEVICE EVALUATION

For our second study, we investigated how participants interacted with the device, and explored their experience using it. They were asked to try out each of the different interactions our device offered as they desired. Due to software issues before running the study, a MIDI keyboard was utilised to manually apply musical effects to the outputted sound. A guitar loop sample from Ableton Live was used as the base audio and effects were added on top of this.

Users were provided with a questionnaire after the study to explore their experiences using the device. This used the 5-point Likert scale to quantify the results. The Likert scale is common when conducting survey research and is very useful in understanding how a particular individual feels about their experience. The survey questions used were adapted from the GEM (Gaming Experience Questionnaire), as this has been shown to be useful form of quantitative research when evaluating musical interfaces [El-Shimy and Cooperstock 2016]. Participants were asked to rank different questions from 1 to 5 on their experience using our device, with 1 representing 'Strongly Disagree,' and 5 representing 'Strongly Agree.' Between these are 'Disagree,' 'Neutral,' and 'Agree.' The results from this study were used to see how effective our device was and gauge overall performance. This information was then visualised using bar graphs. This study was performed with 15 participants.

	1	2	3	4	5
E(X)	3	3	3	3	3
O(X)	0	0	4	10	1

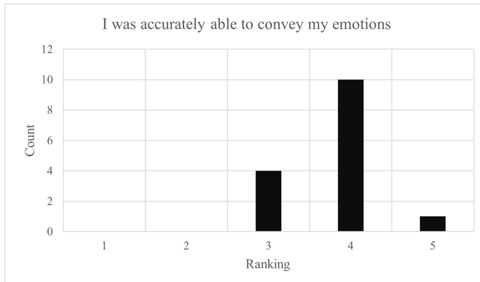

Fig. 5. Results for “I was able to accurately convey my emotions”

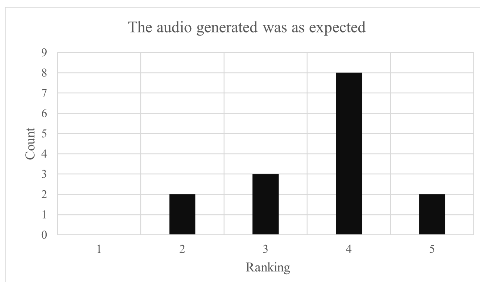

Fig. 6. Results for “The audio generated was as expected”

The results of this study were analysed using a Chi-squared test. Specific questions were chosen to be analysed using this method; others have instead either been analysed using mode and median values - the mean is redundant for analysing a Likert scale - or are depicted as bar graphs in the appendix.

As there is a small sample size (n<30), we have opted to use a significance level of p=0.1. This test uses the results from the question: “I was accurately able to convey my emotions.” The null hypothesis is as follows:

h_0 = the device has no effect on a user’s ability to accurately convey their emotions

The expected value for all ranks is 3. The observed values for each rank are as follows:

The Chi-squared value for this dataset is 24. The critical value for p=0.1, n=15 is 22.307 (Jones, 2020). As the calculated value is larger than the critical value, we can reject the null hypothesis and conclude that the device does influence the user’s ability to accurately convey their emotions. As the mode for the observed values is 4 (‘Agree’), it can be concluded that the device allows users to accurately convey their emotions. The results are shown in Figure 5.

For the question “The audio generated was as expected,” the modal value was 4 (‘Agree’) with 8 votes and the median value was 4 (‘Agree’). This suggests that participants in general believed this statement to be true. The results are shown in Figure 6.

For the question “I enjoyed it,” the modal value was 5 (‘Strongly Agree’) with 8 votes and the median value was 5 (‘Strongly Agree’). This suggests that participants enjoyed their experience. The results are shown in Figure 7.

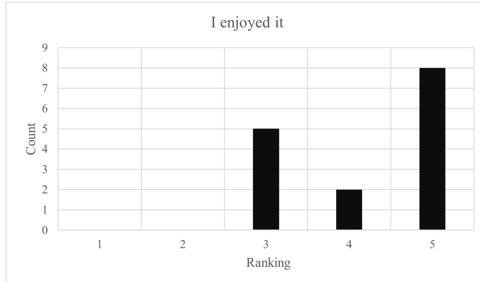

Fig. 7. Results for “I enjoyed it”

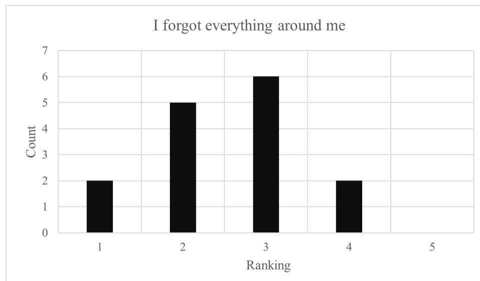

Fig. 8. Results for “I forgot everything around me”

For the question, “I forgot everything around me,” the modal value was 3 (‘Neutral’) with 6 votes and the median value was 3 (‘Neutral’). This does not suggest that users forgot everything around them, and so does not suggest that users entered a ‘flow state.’ If the device is to elicit a flow state for users, further investigation is needed to explore qualities of the device that affect flow, and how they can be improved. The results are shown in Figure 8.

6 DISCUSSION

There are a lot of points of discussion that we want to address about our device. Some relate to the limitations of the device, and some relate to potential alternative uses.

6.1 Limitations

The device is based on both Western models of musical and emotional expression. This limits the device for use outside of the West, as expressions of emotion differ between cultures. One way that these issues might be mitigated is by extending the sensors to cover the entirety of the device’s surface, allowing for the development of alternative interactions in future. For the musical effects, one solution could be to create alternative maps for the circumplex model to cater to various different cultural backgrounds, that can then be interchanged by the user as they see fit. These changes would allow more flexibility in the device for non-Western cultures, and therefore make it more inclusive and accessible.

Another possible limitation of the device is the usage of the circumplex model to represent emotions. Different models of representing emotions, such as the Plutchik Wheel [Semeraro et al. 2021], could be used, and could result in different emotions sharing different musical features due to their organisation. This could also affect the strength of effects, as the circumplex model places all explicitly named emotions around the edge of the model, equidistant from the centre. This is not necessarily the reality, as an individual experiencing anger could have a higher activation than their experience of excitement, but the circumplex model would position both emotions at the same level of the arousal dimension.

6.2 Future Developments

The device has the potential to be used for a variety of uses outside of those we have presented in this document. Some improvements can also be made to increase the functionality of the device, so will be discussed here.

Firstly, the device would benefit from the addition of some sort of visual or haptic feedback when input is successfully recognised. Haptic feedback can increase a user's sense of presence while participating in virtual reality experiences [Gibbs et al. 2022]. Likewise, haptic feedback could be assistive in creating a more realistic experience when interfacing with our device. This would be added to identify when an action has been successfully performed, enabling users to more accurately dial-in the musical effects that they desire.

The device is currently slightly smaller than the average human head, sitting at just under 20cm tall, due to size limitations of the 3D printer used. This resulted in some issues in the device evaluation study, where users accidentally activated multiple contact points at once. Increasing the size of the device to reflect the size of the average person's head would aim to reduce accidental activations, and potentially increase instinctual responses from users.

The software for the device allows for continuous emotional classification, rather than discrete as displayed in the device evaluation section. This feature could not be presented at the evaluation stage, due to issues with Teensy Arduino, but is important to a more nuanced experience of using the device. By classifying emotions, and subsequentially musical effects, users can utilise the device more intuitively, and with more accuracy, and so create an experience more similar to an instrument than a technical device.

The next step to take with the device would be to test how users of different levels of musical ability utilise our interface, using the OMSI (Ollen Music Sophistication Index) [Institute [n. d.]]. This is a tool used to classify musical sophistication within participants. Using this, the ways in which users of different levels of musical ability use the device could be analysed, and any alterations that could be made to increase the device's usability for both inexperienced and experienced musicians could be explored, along with whether those alterations could create benefits over multiple levels of ability.

The use of the device with non-verbal individuals to assist with communication and emotional expression is also important to explore. Musical therapy has been shown to be beneficial to non-verbal, leading to an increase in communication [Silverman 2008]. Therefore, the device, with its unique form factor and links between emotional expression and music, could be beneficial for exploring communication for non-verbal individuals.

7 APPENDIX

7.1 Teamwork and Inclusivity

Our group is made up of people from various nationalities and backgrounds, which was useful in providing different perspectives on the problem at hand. Each of us has different skillsets, knowledge bases and interests. We accounted for this by ensuring each person got a chance to voice their opinions in discussions and designating tasks based on what individuals felt most comfortable doing and where they felt they could provide most value. As a team of male-identifying individuals, we acknowledge that our team may lack insight from female and non-binary identifying people, however, we hope that us undertaking research with a range of diverse participants has removed any potential bias that could have occurred as a result.

7.2 Video Demonstration

https://youtu.be/icv_FFv-A_Q

7.3 Device Evaluation Study Graphs

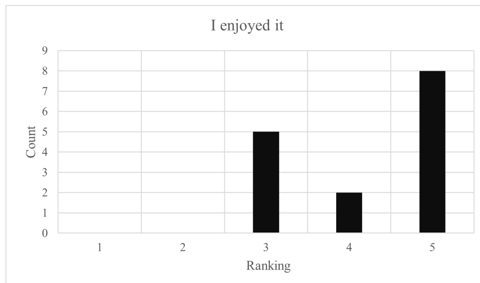

Fig. 9. Results for “I enjoyed it”

Fig. 10. Results for “I felt challenged”

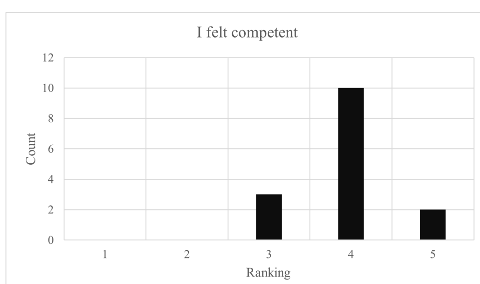

Fig. 11. Results for “I felt competent”

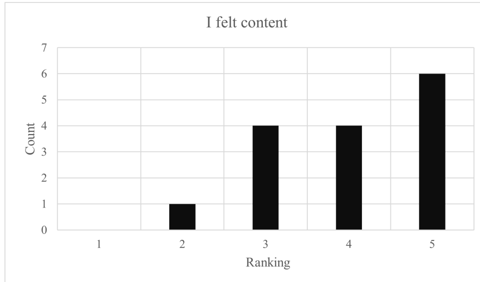

Fig. 12. Results for "I felt content"

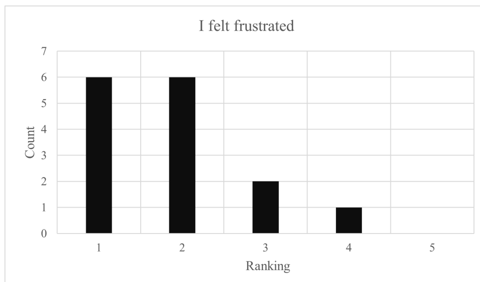

Fig. 13. Results for "I felt frustrated"

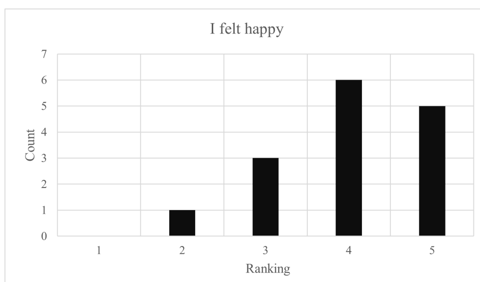

Fig. 14. Results for "I felt happy"

REFERENCES

Arvid J. Bloom and Paula Skutnick-Henley. 2005. Facilitating Flow Experiences Among Musicians. *The American Music Teacher* 54, 5 (2005), 24–28.

Antonio Camurri, Shuji Hashimoto, Matteo Ricchetti, Andrea Ricci, Kenji Suzuki, Riccardo Trocca, and Gualtiero Volpe. 2000. EyesWeb: Toward Gesture and Affect Recognition in Interactive Dance and Music Systems. *Computer Music Journal* 24, 1 (2000), 57–69. http://www.jstor.org/stable/3681851

Chris Creed and Russell Beale. [n. d.]. .

Bhaveek Desai, Benjamin Chen, Sofia Sirocchi, and Kyla A. McMullen. 2018. Mindtrack: Using brain-computer interface to translate emotions into music. *2018 International Conference on Digital Arts, Media and Technology (ICDAMT)* (2018). https://doi.org/10.1109/icdamt.2018.8376491

Matthew Duignan, James Noble, and Robert Biddle. 2010. Abstraction and activity in computer-mediated music production. *Computer Music Journal* 34, 4 (2010), 22–33. https://doi.org/10.1162/comj_a_00023

Dalia El-Shimy and Jeremy R. Cooperstock. 2016. User-driven techniques for the design and evaluation of new musical interfaces. *Computer Music Journal* 40, 2 (2016), 35–46. https://doi.org/10.1162/comj_a_00357

Aalf Gabrielsson and Erik Lindström. 1993. The role of structure in the musical expression of emotions. *Handbook of Music and Emotion: Theory, Research, Applications* (1993), 367–400. https://doi.org/10.1093/acprof:oso/9780199230143.003.0014

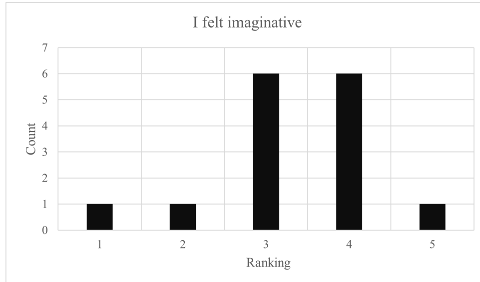

Fig. 15. Results for “I felt imaginative”

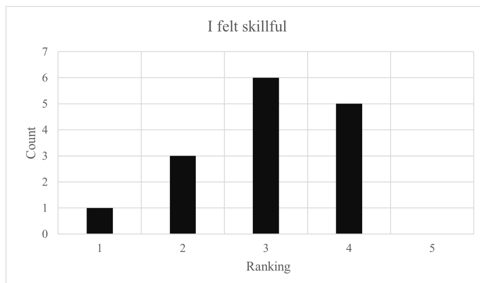

Fig. 16. Results for “I felt skillful”

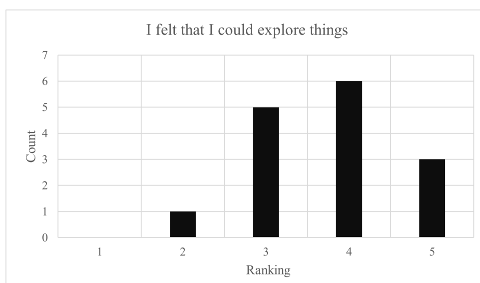

Fig. 17. Results for “I felt that I could explore things”

Janet K. Gibbs, Marco Gillies, and Xueni Pan. 2022. A comparison of the effects of haptic and visual feedback on presence in virtual reality. *International Journal of Human-Computer Studies* 157 (2022), 102717. https://doi.org/10.1016/j.ijhcs.2021.102717

The Marcs Institute. [n. d.]. Ollen Musical Sophistication Index. http://marcs-survey.uws.edu.au/OMSI/omsi.php

janovich. [n. d.]. baseMesh Head (clean). https://www.turbosquid.com/3d-models/free-head-basemesh-clean-3d-model/975205

Sergi Jordà. 2002. FMOL: Toward user-friendly, sophisticated new musical instruments. *Computer Music Journal* 26, 3 (2002), 23–39. https://doi.org/10.1162/014892602320582954

Patrik N. Juslin and Petri Laukka. 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin* 129, 5 (2003), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Steven R. Livingstone, Ralf Muhlberger, Andrew R. Brown, and William F. Thompson. 2010. Changing musical emotion: A computational rule system for modifying score and performance. *Computer Music Journal* 34, 1 (2010), 41–64. https://doi.org/10.1162/comj.2010.34.1.41

A.R. Lopez, António Oliveira, and Amílcar Cardoso. 2010. Real-time emotion-driven music engine. *Proceedings of the International Conference on Computational Creativity, ICCC-10.*

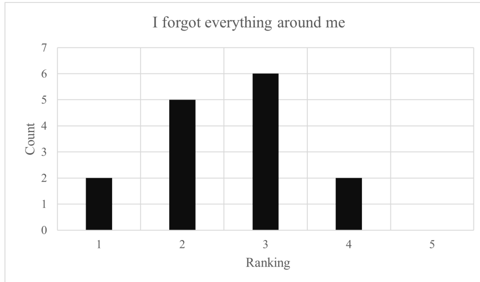

Fig. 18. Results for "I forgot everything around me"

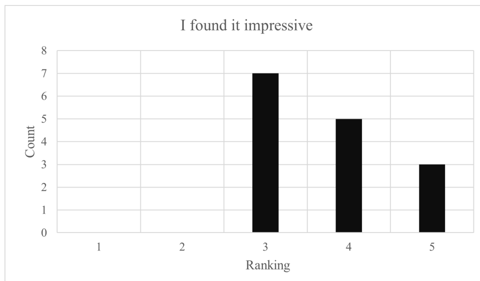

Fig. 19. Results for "I found it impressive"

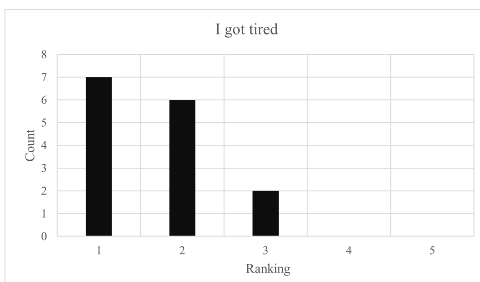

Fig. 20. Results for "I got tired"

Thor Magnusson. 2010. Designing constraints: Composing and performing with Digital Musical Systems. *Computer Music Journal* 34, 4 (2010), 62–73. https://doi.org/10.1162/comj_a_00026

Annaliese Micallef Grimaud and Tuomas Eerola. 2020. EmoteControl: An interactive system for real-time control of emotional expression in music. *Personal and Ubiquitous Computing* 25, 4 (2020), 677–689. https://doi.org/10.1007/s00779-020-01390-7

Jonathan Posner, James A. Russell, and Bradley S. Peterson. 2005. The Circumplex model of affect: An integrative approach to Affective Neuroscience, Cognitive Development, and psychopathology. *Development and Psychopathology* 17, 03 (2005). https://doi.org/10.1017/s0954579405050340

Alfonso Semeraro, Salvatore Vilella, and Giancarlo Ruffo. 2021. Pyplutchik: Visualising and comparing emotion-annotated corpora. *PLOS ONE* 16, 9 (2021). https://doi.org/10.1371/journal.pone.0256503

Michael J. Silverman. 2008. Nonverbal communication, music therapy, and autism: A review of Literature and case example. *Journal of Creativity in Mental Health* 3, 1 (2008), 3–19. https://doi.org/10.1080/15401380801995068

Koray Tahiroğlu, Juan Carlos Vasquez, and Johan Kildal. 2017. Facilitating the musician's engagement with new musical interfaces: Counteractions in music performance. *Computer Music Journal* 41, 2 (2017), 69–82. https://doi.org/10.1162/comj_a_00413

Marc Teyssier. [n. d.]. Skin-On Interfaces. https://marcteyssier.com/projects/skin-on/

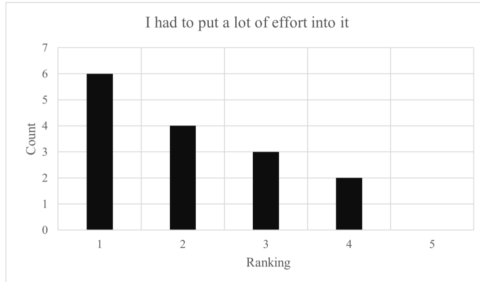

Fig. 21. Results for “I had to put a lot of effort in”

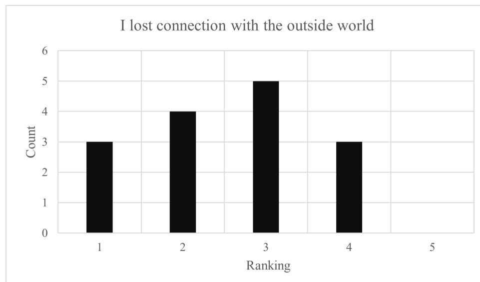

Fig. 22. Results for “I lost connection with the outside world”

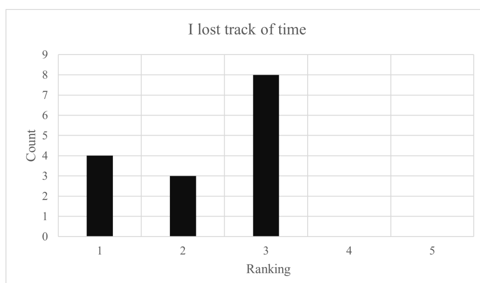

Fig. 23. Results for “I lost track of time”

Francisco Ventura, António Oliveira, and Amílcar Cardoso. 2009. An emotion-driven interactive system. (01 2009).

Marcelo Mortensen Wanderley and Nicola Orio. 2002. Evaluation of input devices for musical expression: Borrowing tools from HCI. *Computer Music Journal* 26, 3 (2002), 62–76. https://doi.org/10.1162/014892602320582981

K.C. Wassermann, Kynan Eng, P.F.M.J. Verschure, and J. Manzolli. 2003. Live soundscape composition based on synthetic emotions. *IEEE Multimedia* 10, 4 (2003), 82–90. https://doi.org/10.1109/mmul.2003.1237553

Todd Winkler. 1995. Making Motion Musical: Gesture Mapping Strategies for Interactive Computer Music. *Proceedings of the 1995 International Computer Music Conference.*

Marcel Zentner, Didier Grandjean, and Klaus R. Scherer. 2008. Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion* 8, 4 (2008), 494–521. https://doi.org/10.1037/1528-3542.8.4.494

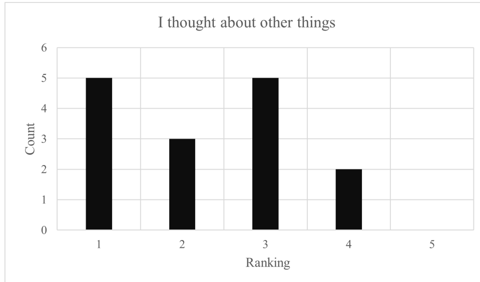

Fig. 24. Results for "I thought about other things"

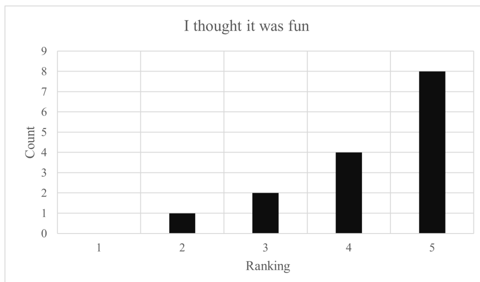

Fig. 25. Results for "I thought it was fun"

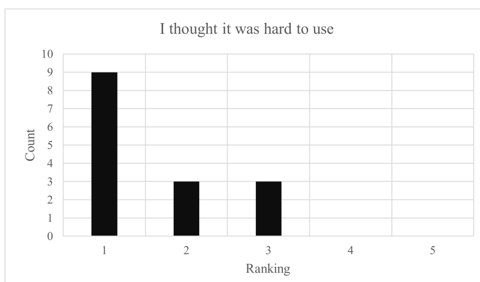

Fig. 26. Results for "I thought it was hard to use"

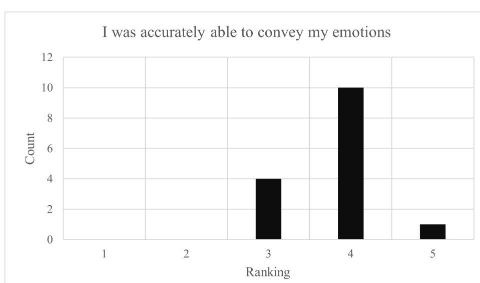

Fig. 27. Results for "I was able to accurately convey my emotions"

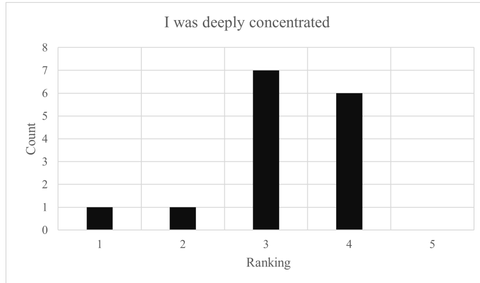

Fig. 28. Results for “I was deeply concentrated”

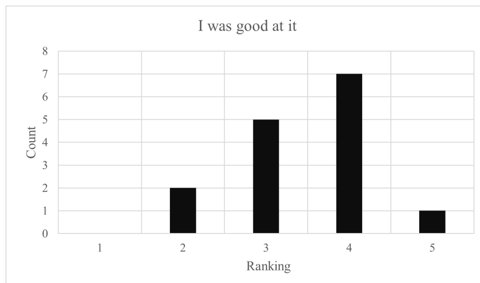

Fig. 29. Results for “I was good at it”

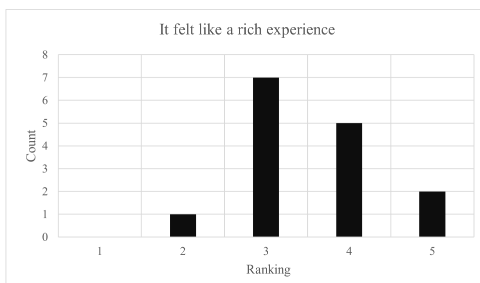

Fig. 30. Results for “It felt like a rich experience”

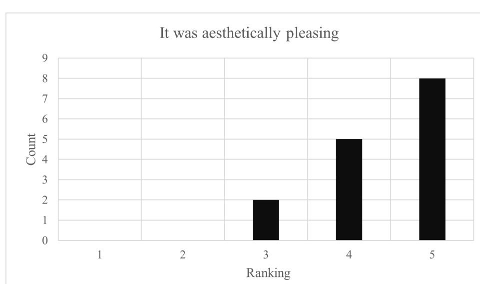

Fig. 31. Results for “It was aesthetically pleasing”

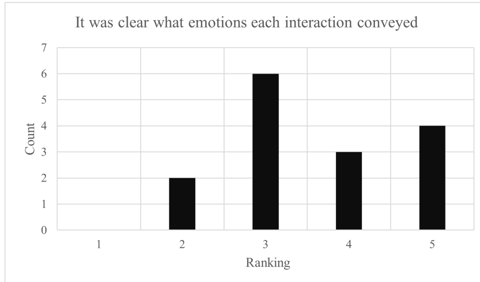

Fig. 32. Results for “It was clear what emotions each interaction conveyed”

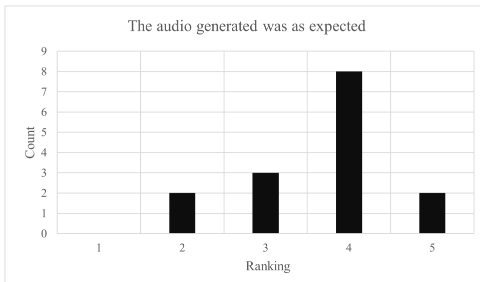

Fig. 33. Results for “the audio generated was as expected”