A Survey on the Use of 2D Touch Interfaces for Musical Expression

Diemo Schwarz, Wanyu Liu, Frédéric Bevilacqua

A Survey on the Use of 2D Touch Interfaces for Musical Expression

Diemo Schwarz
STMS UMR9912
Ircam–CNRS–SU
1, place Igor-Stravinsky
75004 Paris, France
schwarz@ircam.fr

Wanyu Liu
STMS UMR9912
Ircam–CNRS–SU
1, place Igor-Stravinsky
75004 Paris, France
aliu@ircam.fr

Frédéric Bevilacqua
STMS UMR9912
Ircam–CNRS–SU
1, place Igor-Stravinsky
75004 Paris, France
bevilacq@ircam.fr

ABSTRACT

Expressive 2D multi-touch interfaces have in recent years moved from research prototypes to industrial products, from repurposed generic computer input devices to controllers specially designed for musical expression. Practitioners use this type of devices in many different ways, with different gestures and sound synthesis or/and transformation methods. In order to get an overview of existing and desired usages, we launched an on-line survey that collected 37 answers from practitioners in and outside of academic and design communities. In the survey we inquired about the participants' devices, their strengths and weaknesses, the layout of control dimensions, the used gestures and mappings, the synthesis software or hardware, and the use of audio descriptors and machine learning. The results can inform the design of future interfaces, gesture analysis and mapping, and give directions for the need and use of machine learning for user adaptation.

Author Keywords

controllers, touch interfaces, 2D, DMI, NIME, survey

CCS Concepts

•**Applied computing → Sound and music computing;** Performing arts;

1. INTRODUCTION

Touch surfaces are an interface of choice for musical expression. After their debut in the 1990s without pressure sensitivity and only single touch (Korg *Kaoss Pad*, XY-pads in synthesisers), pressure-sensing and multi-touch devices have, after long years of being prototypes (Wessel's *Slabs* [13]) or small-series—rather exclusive—products (Tactex *STC-1000* [8], *Lemur*, *Continuum Fingerboard*), become adopted by the electronic musical instrument industry (Roli *Blocks*, *Linnstrument*, *Joué*), and by small specialist companies (Madrona Labs *Soundplane*, Zvuk Machines *Zvuk9*). In parallel, computer input peripherals (Wacom graphic tablets, PQ-Labs multi-touch screen overlays, Sensel *Morph*) or tablet computers (*iPad*) have appeared and have been used for music performance. These recent industrially produced touch devices have become ubiquitous and really useful, because they are robust and fast, have a high resolution and a low price, and are easily replaceable. At the same time, experimentation is still ongoing with many Kickstarter projects (*TouchKeys*, *Trill*[1]).

This fortunate situation leads many researchers and practitioners to turn to touch interaction for music performance [6], but also poses questions of how they are using these devices, with which gestures, mappings, control layout, and sound generators. For instance, touch surfaces allow for embodied musical performance using control paradigms of timbre spaces [8, 12, 14], symbolic and continuous gestural parameter control, hand-shape recognition, or gesture following and recognition. While these interfaces can be approached intuitively, exploiting their full potential for expressiveness remains an open question. In this regard, the technical implementation, gesture processing and sound mapping, as well as the human gesture learning must be addressed conjointly.

To better understand these questions we launched an online survey[2] that inquired about the participants' devices, their strengths and weaknesses, the layout of control dimensions, the used gestures and mappings, the synthesis software or hardware and the use of audio descriptors and machine learning.

2. PREVIOUS WORK

The recent study by Sullivan and Wanderley [11] can be used as a reference for the demographics of users of general electronic instruments and DMIs, since they went to great lengths to reach non-academic musicians (by posting their survey in music stores and enticing users without a professional motivation to help out fellow researchers/designers by offering a raffle of music gear). Their focus was on factors that contribute to the uptake and continued use of new instruments in performance, but they also offered an overview of other existing questionnaire-based surveys in the NIME field. Our study is complementary as it goes into the details of usage of one specific type of NIME.

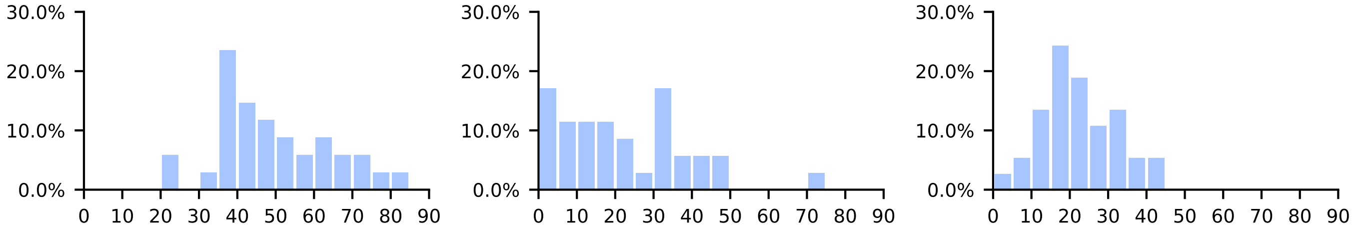
There are studies on specific systems, e.g. Çamcı [2] surveyed 2D touch interaction on a multi-touch tablet in the use case of their customisable granular synthesis software. Regarding machine learning, the link with 2D touch interaction is pursued increasingly often [1, 4, 5].

3. THE SURVEY

We authored an online questionnaire in Google Forms,[2] organised in 8 sections with 46 questions in 3 larger parts.

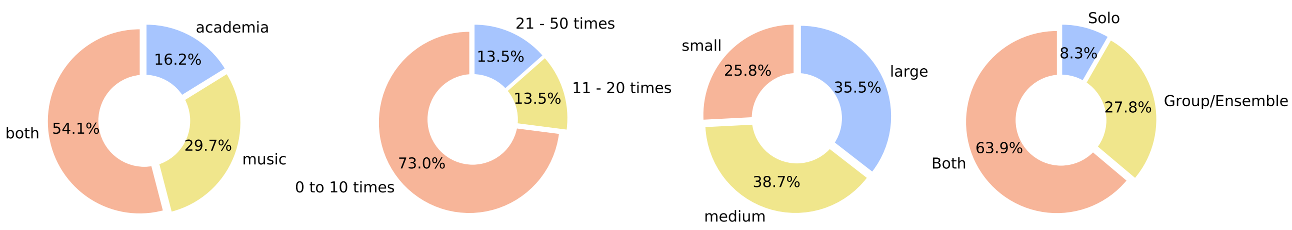
[1]https://www.kickstarter.com/projects/423153472/trill-touch-sensing-for-makers
[2]available at https://forms.gle/PgXSbMPBu73fr4oP6

NIME'20, July 21-25, 2020, Royal Birmingham Conservatoire, Birmingham City University, Birmingham, United Kingdom.



(a) age in years (Q1) (b) years of formal music training (Q6) (c) years of experience in digital audio (Q7)

Figure 1: Distribution of age and musical experience ($n = 34, 35, 37$ responses, respectively).[4]



(a) (b) (c) (d)

Figure 2: (a) Respondents background (Q4, Q5, $n = 37$), (b) number of public performances per year (Q10, $n = 37$), (c) maximum venue size (small ≤ 100, medium ≤ 400) (Q11, $n = 31$), (d) solo/group performance (Q12, $n = 36$).[4]

The first part inquired about the demographics of the participants, the second part about the touch input device and the synthesis software or hardware, the third part about the control layout and the used gestures. For reference, the numbered list of questions is available in the print-ready one-page form.[3]

We sent invitations to relevant mailing-lists in the field of computer music, new interfaces, movement computing, and the communities around major research centres and software environments, and received 37 responses. The questions were always optional and mostly free-form and were analyzed qualitatively using techniques taken from Grounded Theory [10]: Answers underwent one or two rounds of coding to allow qualitative analysis and to uncover common topics and their frequencies of occurrences.

3.1 Demographics

The participants' origin (Q3) was almost exclusively from Europe and North America (the only exception was one participant from China), with France, USA, Canada, UK, Germany making up 3/4 of the origins; 89% of them identified themselves as male, 3 as women, one preferred not to say (Q2). Their age distribution (Q1) is shown in figure 1a, with mean of 48.94 and median of 46 years.

Answers to the questions about background (Q5) and occupation (Q4) in figure 2a were coded for "academia" (code for researchers, teachers, designers, students), "music" (code for professional musicians), and "both".

Regarding musical training, participants had a mean of 21.06 and median of 19 years of formal musical training (Q6, figure 1b), and mean of 20.73 and median of 20 years of experience in digital audio (Q7, figure 1c).

The main genres (Q8) of music the participants perform are listed in figure 3. The list of options is taken from the AllMusic online database,[5] but adapted to the specificities of our target group (similar to [11]) by dividing *electronic* into *EDM*, *Electro-Acoustic*, and adding *Contemporary* to the *Avant-garde/Experimental* category (abbreviated here as *Avant./Expe./Contemp.*). Respondents could choose up to three genres and could specify additional sub-genres or styles (Q9). The latter didn't add much information and is thus not reported here. The number of self-reported public performances per year (Q10), maximum venue size (Q11), and solo or group performance (Q12) is shown in figures 2b–2d.

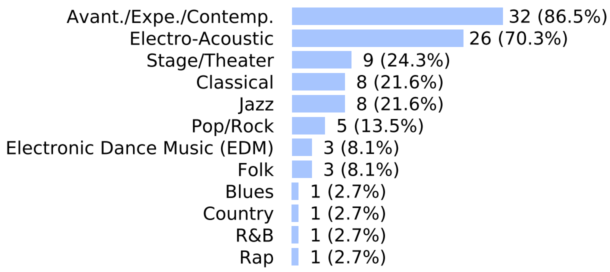


Figure 3: Primary genre(s) of music performed or produced (Q8, $n = 37$).[4]

The panel of our participants are mostly comparable with Sullivan and Wanderley's [11], except for number of performances per year, where they report 12% "50 times or more", and gender, where they had 29% non-male responses, probably reflecting a more diverse panel of potential and actual respondents, extending further into non-academic musicians and being less specific than our target group.

3.2 Information about the Input Device

The actual and planned device use (Q14) is shown in figure 5, the device names (Q15) and shape (Q16) in figure 6. We see here a large variety of devices used, with the most ubiquitous and longest existing ones coming up top: *iPad*, computer trackpad, touchscreens, Wacom tablet, followed by recent specialised devices Sensel *Morph*,[6] Roli *Lightpad*

[3] Annex to https://hal.archives-ouvertes.fr/hal-02557522
[4] Note that we report all percentages relative to the number of participants n who answered the respective question.
[5] http://www.allmusic.com/genres
[6] https://sensel.com/pages/the-sensel-morph

Figure 4: Examples of commercial devices used by the participants. From top left to bottom right: *Morph*, *Lightpad* Block, *Touché*, *Linnstrument*, *Continuum*, *Joué*, *Kaoss Pad*, and *Soundplane*. Pictures taken from the respective official websites.

Block,[7] Expressivee *Touché*.[8] Not shown are single mentions of other devices like Roger Linn Design *Linnstrument*,[9] Korg *Kaoss Pad*,[10] Madrona Labs *Soundplane*,[11] Haken Audio *Continuum*,[12] *Joué*.[13]

The distribution of the physical size of the largest side in cm (Q17) is shown in figure 8, with a mean of 29.21, median of 24, standard deviation of 22.95, min of 7, max of 120.

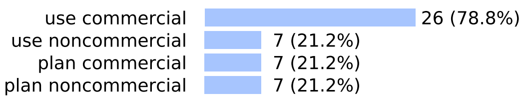


Figure 5: Device use (Q14, $n = 33$).[4]

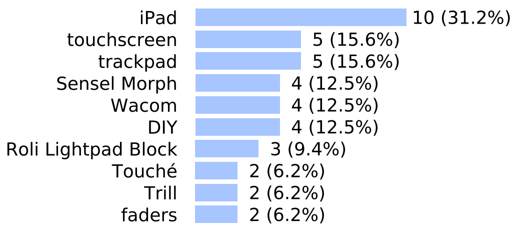


Figure 6: Device name (Q15, $n = 32$).[4]

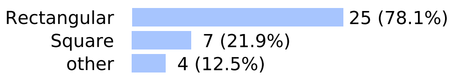


Figure 7: Device shape (Q16, $n = 32$).[4]

The distribution of one- or two-handed interaction and used body part or object (Q18) is shown in figures 9 and 10 (here, "others" could mean hand palm, arms, etc.). Interestingly, two-handed use is also reported for small device sizes, but above 30 cm, it is exclusive.

Finally, we asked about what the participants liked the most (Q20) and the least (Q21) about their devices. The mentioned features were coded into topics and topic groups

[7] https://roli.com/products/blocks/lightpad-block
[8] https://www.expressivee.com/touche
[9] https://rogerlinndesign.com/linnstrument
[10] https://wikipedia.org/wiki/Kaoss_Pad
[11] https://madronalabs.com/soundplane
[12] https://hakenaudio.com/continuum-fingerboard
[13] https://play-joue.com

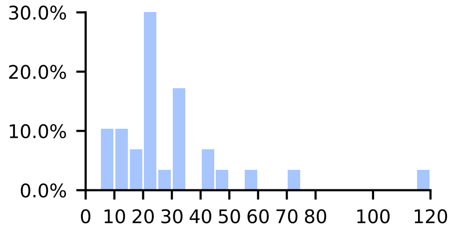


Figure 8: Device size in cm (Q17, $n = 29$).[4]

Figure 9: Hand usage (Q18, $n = 29$).[4]

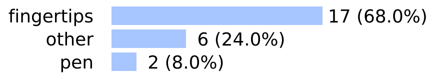


Figure 10: Touching body part or object (Q18, $n = 25$).[4]

Feature Class	Liked		Disliked	
	Topic Group	Feature Topic	Topic Group	Feature Topic
a-priori device feature	portable	portability	small size	small size
		small size		
		light weight		
	available	availability		
functional capabilities	versatile	customizability	unreliable	unreliable
		versatility		
	ease of use	ease of use	difficult setup	difficult setup
		bundled synth		
interaction capabilities	nuance	sensitivity	low nuance	low sensitivity
		pressure		no pressure
		resolution		low resolution
	multitouch	multitouch		
	feedback	tactile feedback		
		screen		

Figure 11: Liked and disliked features of the device by class and topic group.

shown in figure 11, and fall into three classes coded by colour in the following.

In both liked and disliked individual topics in figures 12 and 14, *ease of use* vs. *difficult setup* came on 2^{nd} and 1^{st} place, respectively, stressing the practical requirements for live performance. However, in the grouped topics in figures 13 and 15, among the top three liked topic groups are *nuance*, *ease of use*, matching the top disliked *low nu-*

ance, difficult setup. Regarding *customizability*, 3 occurrences were references to hardware customizability, e.g. for the *Trill* touch sensors, the others to software.

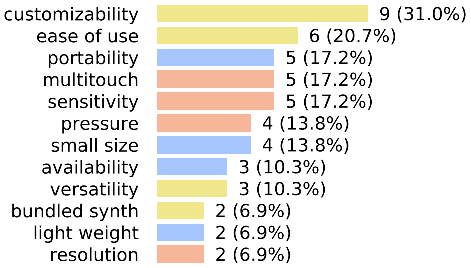


Figure 12: Most liked device feature topics with occurrences greater than one (Q20, $n = 29$).[4] Colours correspond to the feature classes defined in figure 11.

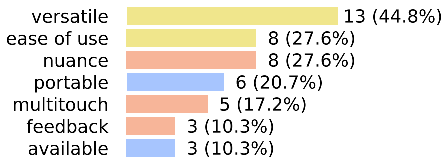


Figure 13: Most liked device feature topic groups with occurrences greater than one (Q20, $n = 29$).[4] Colours correspond to the feature classes defined in figure 11.

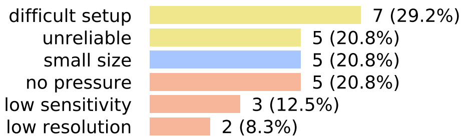


Figure 14: Least liked device feature topics with occurrences greater than one (Q21, $n = 24$).[4] Colours correspond to the feature classes defined in figure 11.

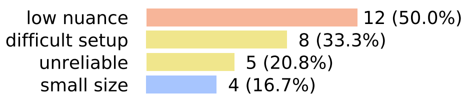


Figure 15: Least liked device feature topic groups with occurrences greater than one (Q21, $n = 24$).[4] Colours correspond to the feature classes defined in figure 11.

3.3 Synthesis Software/Hardware

In this section, we inquired about the synthesis software or hardware used with the touch device. Multiple responses were possible. We see in figure 16 an almost pervasive use of programmable interactive environments like Max, PureData, or Supercollider, but also a higher than expected part of hardware synthesisers (20 mentions). We also asked about the used communication protocol (Q19), if known, and collected 4 mentions of OSC, 3 of MIDI (out of 11).

We then asked if audio descriptors were used (Q27), as with corpus-based concatenative synthesis [7, 9], for instance, which descriptors were used (Q28), and what additional descriptors were suggested as useful (Q29). This last question yielded interesting answers stressing high-level musical features, derived, textural and spatial descriptors.

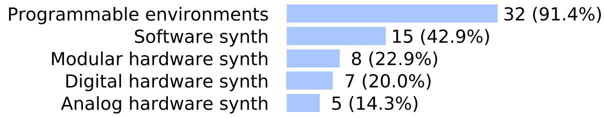


Figure 16: Topics with occurrences greater than one for questions about type of synthesis software or hardware (Q22, $n = 35$).[4]

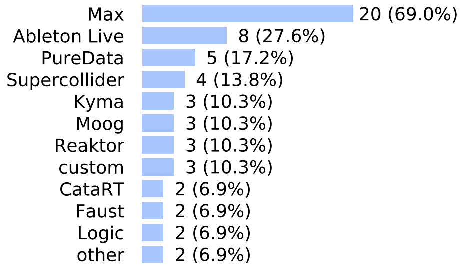


Figure 17: Topics with occurrences greater than one for questions about type of synthesis product (Q23, $n = 29$).[4]

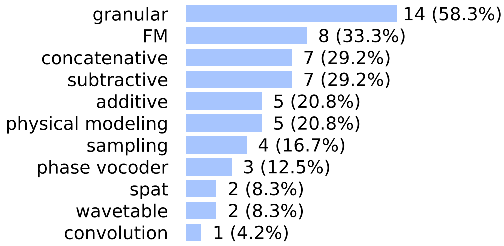


Figure 18: Mentions of synthesis methods (Q25, $n = 24$).[4]

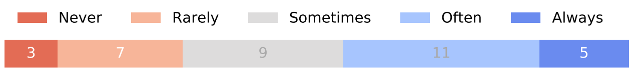


Figure 19: Ratings for "Descriptors are useful" (Q27, $n = 35$).[4]

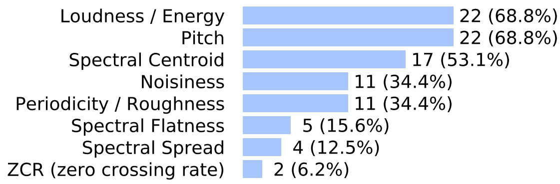


Figure 20: Use of descriptors (Q28, $n = 32$).[4]

3.4 Layout

In this section, we asked about the layout (positioning) of sounds/parameters on the devices' 2D touch surface. First of all, 58% of respondents use one 2D space, but 36% use more than two or a variable number during a performance (Q30), figure 21. The responses for continuous axes or discrete area based layout (Q32) and manual or automatic layout (Q33) are given in figures 22 and 23. Regarding automatic methods, descriptors and t-SNE are mentioned.

3.5 Gestures for Control of Audio Synthesis

Here we asked about the type and character of gestures or movements that are used on the touch device(s).

An overwhelming part of the respondents find gestural interaction useful (Q35, figure 24). When asked to de-

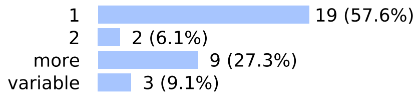


Figure 21: Number of spaces (Q30, $n = 33$).[4]

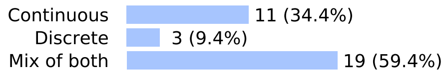


Figure 22: Types of layout (Q32, $n = 32$).[4]

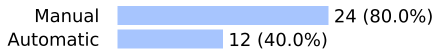


Figure 23: Layout method (Q33, $n = 30$).[4]

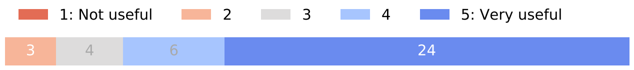


Figure 24: Rating for "Using gestures to control synthesis" (Q35, $n = 37$).[4]

scribe the gestures they are using, the participants came up with topics pertaining to their character (*continuous, discontinuous, slow*), size (*micro, meso, macro*), execution (*tap, swipe, press*), and one metaphorical description: *instrumental* (Q36, figure 25).

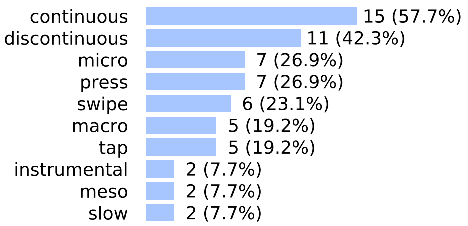


Figure 25: Topics in description of gestures (Q36, $n = 26$).[4]

For the following 11 Likert-scale questions (Q38, figure 26) on the influence of gesture use on the creative process, we can see high ratings for the first group of creativity-related questions (1–4), but less high ratings for the second group of efficiency-related ones (5–8). Especially, the results of the last group of questions on gesture learning (9–11) can be interpreted that for about half the respondents, gestures are easy to learn, but not so easy to teach to others, but definitely lead to expert or virtuoso playing. This hints at a highly individual approach to gestural control of performance, where each practitioner improves herself, with more difficulty to teach others her art and replicate the gained knowledge.

3.6 Use of Machine Learning

At the end of the gesture section, we inquired about the use of any type of machine learning techniques to control sound (Q39). Of the 28 respondents, 21.4% answered positively, 78.6% negatively. The only mentioned techniques were kNN (k-nearest neighbour matching) and NMF (non-negative matrix factorisation). In the follow-up question how the participants could imagine using machine learning in the future (Q40), the most prominent topic were gestures and generativity, automatic 2D layout and spatialisation, but several respondents also mentioned technical limitations they were encountering in machine learning, or said outright they were not interested.

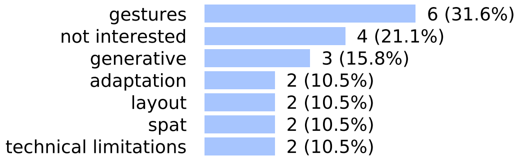


Figure 27: Topics for imagined use of machine learning (Q40, $n = 19$).[4]

The results of these two questions hint at a too high informational and technical barrier for the musicians and performing academics who participated in the study to take up machine learning, and some clearly express an attitude against machine learning use for music performance: "No - I (or the performer) will control them, thank you."

3.7 Desired Device and Gestures

Finally, we asked about the participants' imagined ideal 2D touch input device and how they would like to use it for control of sound synthesis or transformation.

The topics occurring in the "dream device" are given in figure 28. They mention mainly technological improvements like having a screen under or as the touch surface, its material, haptic feedback, a bigger size, and better resolution and data rate. Only some propositions go further, proposing flexible devices, different shapes, extensions to 3D, and transparency: "It would be large and transparent, that the audience could see through it".

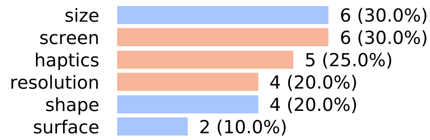


Figure 28: Topics for the desired dream device (Q41, $n = 20$).[4] Colours correspond to the feature classes defined in figure 11.

Together with the responses to the accompanying gestures and synthesis devices, we can observe that a large part of respondents are satisfied with their own device, or desire only incremental improvements. This is interesting, as 3/4 of them are academic researchers/designers/teachers who should at least know, or even develop themselves prototypes of new devices. However, for their practical musical use, they would prefer their tried and tested instruments, if they were just a little bigger. This might also hint at the interfaces transitioning from being seen as prototypical controllers towards actual musical instruments, into which time needs to be invested to master them [3], thus needing them to be mostly unchanging.

4. CONCLUSIONS

This survey showed a sample of usages of 2D touch interaction for musical expression. An important limitation of the study lies on expected biases in the representation of the respondents, and the relatively small size of the group. Therefore, this study should be completed by further investigations, in particular to better target currently underrepresented groups in our study (for example the number of female respondents is very small, and likely points towards biases in the method used to reach respondents). Also, different methodologies such as semi-structured interviews could complement different perspectives. Nevertheless, this study represents an important first step by validating the

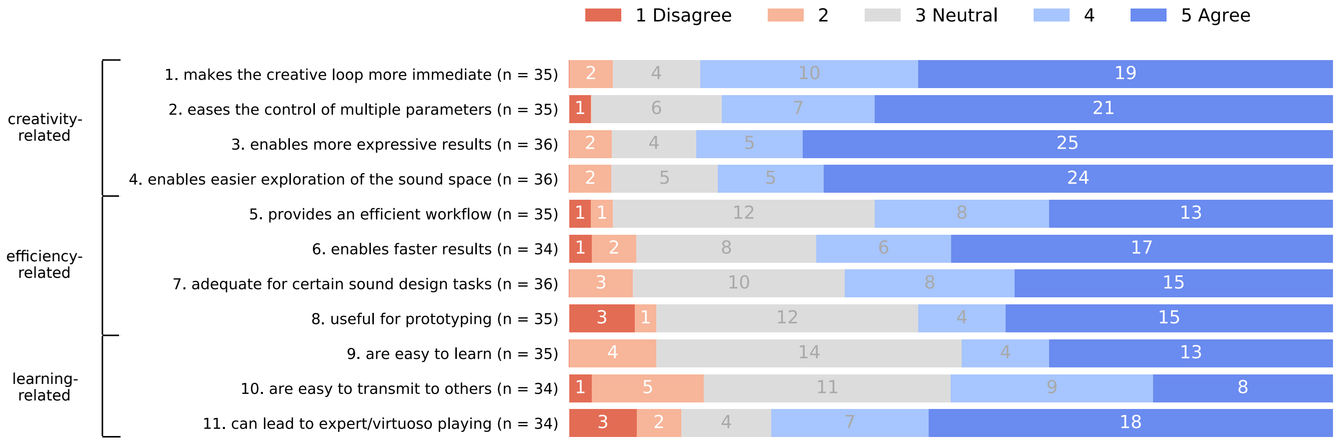


Figure 26: Ratings for Likert-scale questions on using gestures for real-time control of audio synthesis (Q38).[4]

relevance of most of the questions, leading to new insights about this emerging expressive practice.

We do get the following interesting results: First, many participants favoured easily available, well established general computing devices (tablet computers, trackpad, graphics tablets) over devices specially developed for musical expression (Sensel, Roli, Touché), accepting the loss of some expressive input dimensions (pressure, multi-touch). This also shows in the most and least liked features (*ease of use, portable*, and *difficult setup, unreliable*, respectively), but is here balanced by desire for versatility and nuance. This result leads us to predict a future drift towards more use of specialised devices as these become better established. Somehow unexpected was the high use of two-handed interaction, even on the smaller devices. Regarding synthesis methods, concatenative synthesis is now well-established in 3rd place behind granular and FM synthesis, and audio descriptors are used often or always by almost half the participants. In the gesture-related questions, we found a clear adherence to their usefulness for musical expression and creativity, while recognising difficulties in their learnability and teachability. This question should be further examined. Finally, the forward-looking questions can inform the design of future devices and gesture analysis software, and brought to light a gap in the actual practical use of machine learning for expressive interaction, a gap that might easily be overlooked given the hypeful over-presence of trendy machine learning in publications and social media.

5. ACKNOWLEDGMENTS

First of all, we would like to thank the 37 study participants for their valuable time and their very rich contributions, and John Sullivan for reaching out to collected contacts. This work is partially funded by the Agence Nationale de la Recherche grant ANR-18-CE33-0002 "Enabling Learnability in Embodied Movement Interaction" (ELEMENT).

6. ETHICAL STANDARDS

Participants were clearly informed about the aims of the study, were asked explicit confirmation of consent to participate, and had the right to decline participation and withdraw from the study at any time without further consequences. The data was recorded, analysed, and used for developmental and publication purposes, with complete respect for confidentiality. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7. REFERENCES

[1] C. Bascou, V. Emiya, and M. Laurière. The problem of musical gesture continuation and a baseline system. In *International Computer Music Conference (ICMC)*, Utrecht, Netherlands, Sept. 2016.

[2] A. Çamcı. GrainTrain: A hand-drawn multi-touch interface for granular synthesis. In *New Interfaces for Musical Expression (NIME)*, pages 156–161, 2018.

[3] J. Cannon and S. Favilla. The investment of play: expression and affordances in digital musical instrument design. In *International Computer Music Conference (ICMC)*, 2012.

[4] L. Hantrakul and Z. Kondak. GestureRNN: A neural gesture system for the roli lightpad block. In *New Interfaces for Musical Expression (NIME)*, pages 132–137, 2018.

[5] C. Martin and J. Torresen. An interactive musical prediction system with mixture density recurrent neural networks. In *New Interfaces for Musical Expression (NIME)*, June 2019.

[6] P. Nyboer. Experiments in pressure sensitive multi-touch interfaces. NIME workshop, June 2018.

[7] D. Schwarz. Corpus-based concatenative synthesis. *IEEE Signal Processing Magazine*, 24(2), Mar. 2007.

[8] D. Schwarz. The sound space as musical instrument: Playing corpus-based concatenative synthesis. In *New Interfaces for Musical Expression (NIME)*, pages 250–253, Ann Arbor, MI, USA, May 2012.

[9] D. Schwarz, G. Beller, B. Verbrugghe, and S. Britton. Real-Time Corpus-Based Concatenative Synthesis with CataRT. In *Digital Audio Effects (DAFx)*, Montreal, Canada, Sept. 2006.

[10] A. Strauss and J. Corbin. Grounded theory methodology. *Handbook of qualitative research*, 17:273–85, 1994.

[11] J. Sullivan and M. M. Wanderley. Surveying digital musical instrument use across diverse communities of practice. In *Computer Music Multidisciplinary Research (CMMR)*, page 745, 2019.

[12] D. Wessel. Timbre space as a musical control structure. *Computer Music Journal*, 3(2):45–52, 1979.

[13] D. Wessel, R. Avizienis, A. Freed, and M. Wright. A force sensitive multi-touch array supporting multiple 2-d musical control structures. In *New Interfaces for Musical Expression (NIME)*, pages 41–45, 2007.

[14] D. Wessel and M. Wright. Problems and prospects for intimate musical control of computers. *Computer Music Journal*, 26, 12 2001.