A Topo-Phonic Table for Tangible Sonic Interaction

Diemo Schwarz
UMR STMS IRCAM–CNRS–UPMC
`schwarz@ircam.fr`

Grégoire Lorieux
IRCAM
`lorieux@ircam.fr`

Emmanuelle Lizère
composer
`em.lizere@gmail.com`

Ariadna Alsina Tarrès
composer
`ariadnaalsina@gmail.com`

Frédéric Bevilacqua
UMR STMS IRCAM–CNRS–UPMC
`bevilacqua@ircam.fr`

ABSTRACT

We describe an interactive sound installation project consisting of a specially designed table with a patterned surface that allows rich and expressive tangible sonic interaction by rubbing, scraping and hitting it with the hands or objects. We will concentrate on the aspects of gesture tracking and analysis in order to detect the location and type of interaction on the surface, although only a single piezo signal is the sole source of information. For this we have recorded a database of varied types of interactions and report satisfactory classification accuracy using Gaussian mixture models on an MFCC representation of the input. Sound synthesis is performed by convolution of the piezo signal with a set of impulse responses mixed according to the GMM output, with the additional feature of triggering of sounds by attack detection.

1. INTRODUCTION

We give an account of an interactive installation project called *Café Topo-Phonie* at the intersection of art, design, and technology, aimed primarily at children. We will concentrate on the aspects of gesture tracking and analysis in order to detect the location and type of interaction on a surface, used to drive expressive and varied sound synthesis. The installation is part of the exhibition *Le son au bout des doigts* at the Centre Pompidou including 3 interactive installations [1] that relate a specific type of audio–graphic interaction and sound world designed by composers Ariadna Alsina and Emmanuelle Lizère with works of the collection of the Pompidou's contemporary art museum.

The installation has the form of a wooden table and 6 seats, designed by Jakob+MacFarlane design studio (an habitual collaborator of the CGP, having designed the interior of its *Georges* rooftop restaurant). The surface of the table is sculpted with patterns made of CNC-machined incisions and ridges that allow to create rhythmic or continuous impulses when scraping over them with the hands or the provided plates and cutlery.

The interaction is solely picked up by piezo contact microphones, but the artistic aim is to generate varied sounds depending on the object that is used (hands, cutlery, plates), the type of interaction (rubbing, scratching, hitting), and, most of all, the place of interaction (in the centre or at the borders), hence the name *topo-phonic* table. These aims pose a tough problem for sound and gesture analysis, and we'll see how we tackled them to achieve a suffienctly precise tracking for rich and meaningful interaction (that allows the mediators to generate a half-hour narrative related to the artworks) under the constraints of fabrication of the tables that should be robust enough to withstand the assault of thousands of children and then travel to other cities.

Our solution is to classify the spectrum of the contact interaction sound with a Gaussian mixture model (GMM), in order to distinguish type and place of the interaction (section 4), to determine which sound is used as impulse response for convolution with the input (section 5). Additionally, we detect attacks to play back the impulse response or other "surprise" sounds (section 6).

2. RELATED WORK

Simultaneous contact gesture classification and use with resonators have been introduced by the ISMM team [3, 4, 5], that lead to the *MO* modular musical objects [2]. The MO software introduces gesture recognition to distinguish different contact gestures (fingertip or -nail scratching, for instance) to then drive different resonators (physical models of strings).

Based on this work is the *Mogees* project [3] [6], which excites physical models of string or bell resonators running on mobile devices with the input of a specially built contact microphone, allowing to hit, scratch, and strum any surface and turn it into a musical instrument.

We leveraged the use of arbitrary sounds as impulse response (IR) for convolution in combination with a corpus-based approach to choose the IR from a large collection of sounds by high-level audio descriptors in the "Rich Contacts" work [7].

Independently of this research, the first author has been using piezo pickups since 2009 on various surfaces, that allow to hit, scratch, and strum a corpus of sound [8], exploiting its nuances according to the sound of the impacts

---

[1] The two others are Dirty Tangible Interfaces (DIRTI) [1, 2] with new interaction materials such as feathers, tree bark, and a voice-controlled drawing installation called MOC by Lab 212.

[2] See http://youtu.be/Uhps_U2E9OM?t=1m7s at 1:07.

[3] http://www.brunozamborlin.com/mogees/

which is analysed and mapped to the 2D navigation space of the CATART software [4].

In parallel, Puckette [9] has proposed the use of piezo-captured percussive performance as excitors of nonlinear reverberators in 2011, with pre-processing of the piezo in order to remove the resonances of the physical system. The paper also makes explicit what is so interesting in keeping the audio signal from the exciter, by opposition to commercial trigger pickups: "*for instance, sliding a brush over a drum trigger isn't likely to produce anything useful, whereas doing the same thing on an instrument that operates directly on the audio signal from the contact microphone (as we do here) has the possibility to create a wide range of useful musical sounds*".

Last, Tomás [10, 11] uses specially engraved wooden panes as controllers to drive corpus-based synthesis in an instrumental or installation context.

3. DESIGN ASPECTS AND CONSTRAINTS

Each of the 3 tables has to have place for 6 children. The surface is sculpted with a non-regular pattern within which 6 areas for plate and cup are marked (see figure 1).

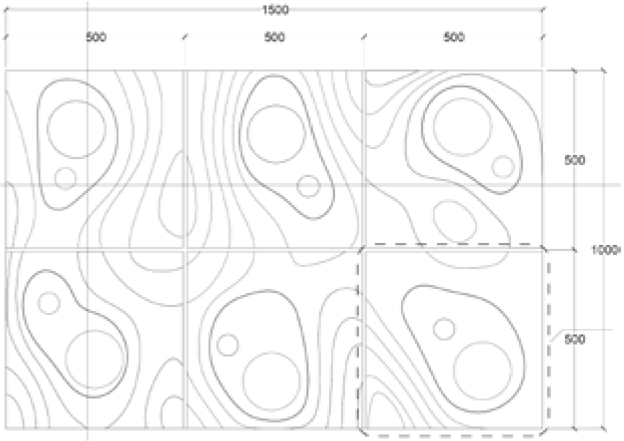


Figure 1. Top view of first design of table surface (Jakob+MacFarlane).

This design was realised as a prototype shown in figure 2 in MDF with rather large structures, which prompted the suggestion to create a more varied structuring of the surface shown in figure 3, exhibiting smaller and varied ridge distances to obtain different sustained tones when scratching an object over the surfaces. We also proposed the idea of creating rhythmic structures by alternating dense and less dense ridge patterns along a path (figure 4).

The final design, prototype, and finished tables are shown in figures 5, 6, and 12 where clearly the idea of a landscape that one can traverse in search of different sounds is referenced.

It shows a more organic patterning and one of the main features of the design: the separate inner pane to place cup and plate within the outer pane. This was intended to make it easier to distinguish the place of interaction, and to produce a specific sound response when it takes place at the "proper" position for cup and plate, as opposed to an outside "disorderly" placement. Two separate piezos, one under the inner pane, one under the outer pane, were used to pick up the interaction in the centre and outside, but because of constraints of number of audio channels, each pair of 2 piezos were mixed to one channel in a 16-in to 8-out DI box. The challenge is now to nevertheless distinguish place and type of interaction.

Figure 2. Photo of first prototype of table surface.

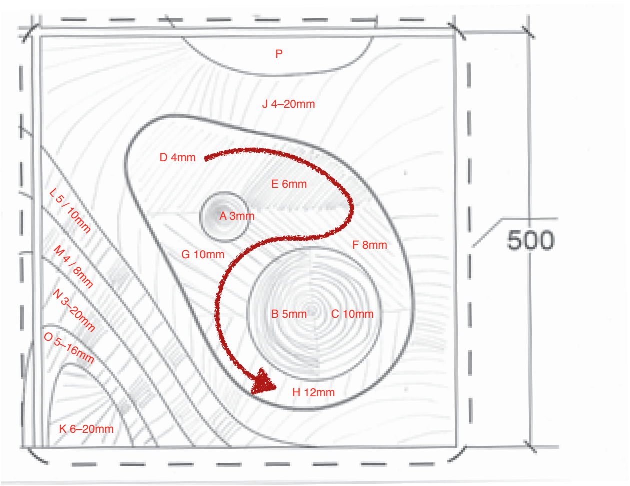


Figure 3. Proposal for musical surface structuring.

4. GESTURE CLASSIFICATION BY GMMS

The interaction sound captured by the 2 piezos for each plateau is transformed into a 12-band MFCC spectrum representation with window size of 46 ms and hop size 12 ms. The stream of MFCCs is then smoothed with a moving average filter of order 20 (corresponding to a window of 232 ms) and sent through a pre-trained GMM decoder. The likelihoods from GMM classification for all classes are sent to the synthesis module (see section 5).

Training is done by recording the MFCC representation of examples of all positions, interaction types and objects that we wish to distinguish into a track of a MuBu multi-buffer container [12] [5]. These recordings and their labels are used to train the GMM. The MFCC preprocessing is implemented by the PiPo extensions [6] for audio and data stream processing in MAX/MSP [13]. Training and decoding are implemented by the `mubu.gmm` external [14].

---

[4] http://ismm.ircam.fr/catart

[5] http://ismm.ircam.fr/mubu

[6] http://ismm.ircam.fr/pipo

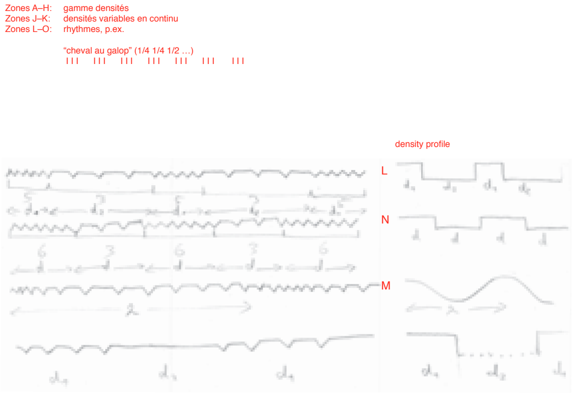


Figure 4. Proposal for musical patterns engraved in the surface structuring.

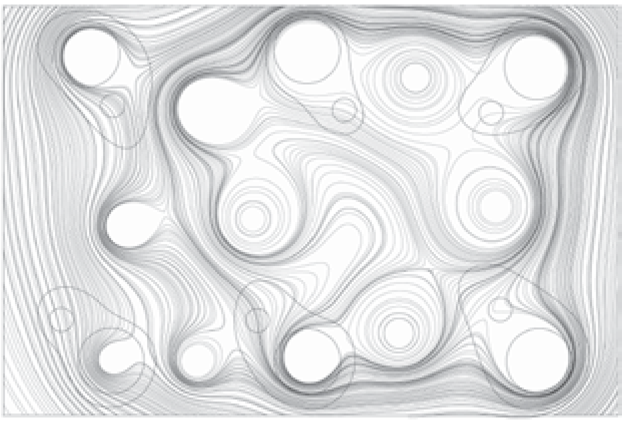


Figure 5. Second "topographic" surface design (Jakob+MacFarlane).

4.1 Classification Accuracy

In order to assess the possibilities and limits of discrimination of gestures, we recorded a database of many different types of interactions to test the accuracy of the classification. The database is organised according to these categories: Each interaction takes *part* either on the inner or outer pane, has a specific *position* on the pane (big or small end for the inner pane, 4 corners for the outer), and performs an *action* (rub, scratch, tap, drum) using a *tool* (hand, fingernails, cutlery) on an *object* (none, i.e. the table itself, plate, cup). Table 1 gives a summary of categories and classes. Each example recording lasts for around 10 s.

Note that these categories and classes are much more detailed than is necessary and useful for the installation. The aim was to highlight what kinds of interactions could possibly be distinguised to inform the sonic interaction design decisions, that will assign different types of sounds and thus artistic meaning to only some of the interactions in the test database.

We trained the system on the database with the first second of each example removed, and ran a decoding test on that first second. The results of the classification accuracies in terms of individual MFCC frames are given in figures 7–10. We can see that all categories except *position* are rather well identified, with 84% and 92% of the MFCC frames correctly classified for the two classes of the *part* category, 92% of *tool* except *hand* with only 60%, $\geq$ 69% of *object*, $\geq$ 74% of *action*.

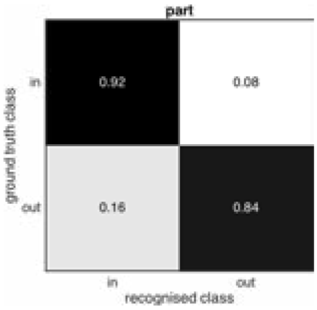
Figure 6. Second prototype of table surface.

Category	Classes	Number of Examples
part	in	34
	out	68
position	big	17
	small	17
	bottomleft	17
	bottomright	17
	topleft	17
	topright	17
tool	cutlery	48
	hand	42
	nails	12
object	cup	24
	none	42
	plate	36
action	drum	6
	rub	42
	scratch	24
	tap	30
Total		**102**

Table 1. Summary of database contents.

5. AUDIO SYNTHESIS BY CONVOLUTION

In the installation, the mixed sound of the two piezo pickups on the central and surrounding pane of each of the 6 places of the each table are sent to a Max/MSP patch that performs the same MFCC and smoothing analysis as above and sends this data stream to the `mubu.gmm` decoder. The decoder outputs a list of likelihoods, one for each of n trained label. One label corresponds here to one type of interaction. The likelihoods are mapped to amplification factors, that are applied to n copies of the piezo audio signal. Each copy is then convolved with a different impulse response, specifically design for that type of

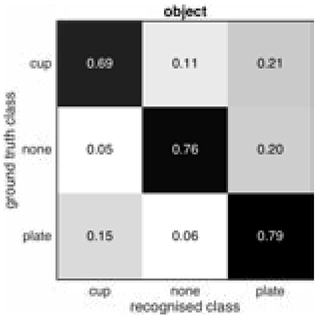


	in	out
in	0.92	0.08
out	0.16	0.84

Figure 7. Confusion matrix for *part* classification results.

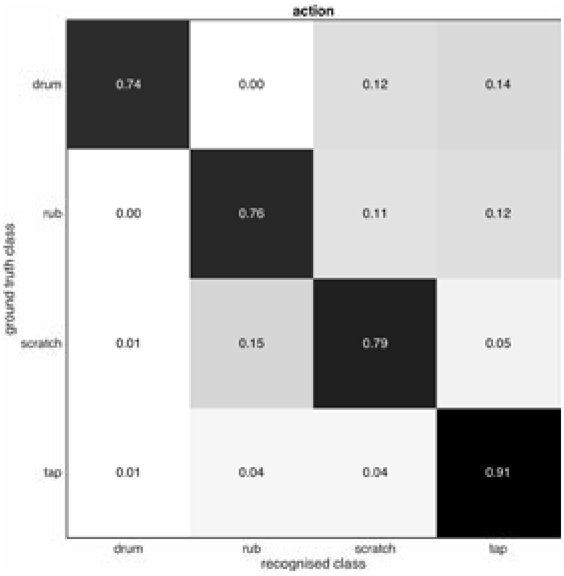
	cutlery	hand	nails
cutlery	0.92	0.07	0.01
hand	0.35	0.60	0.06
nails	0.03	0.04	0.93

Figure 8. Confusion matrix for *tool* classification results.



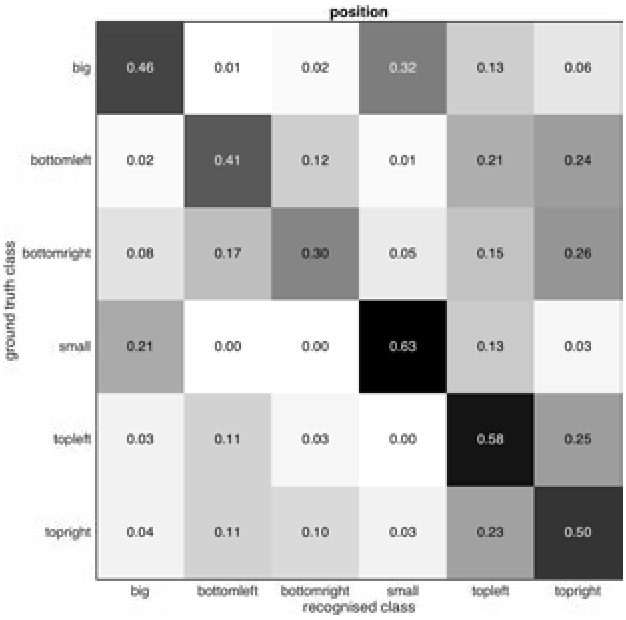
	cup	none	plate
cup	0.69	0.11	0.21
none	0.05	0.76	0.20
plate	0.15	0.06	0.79

Figure 9. Confusion matrix for *object* classification results.



	drum	rub	scratch	tap
drum	0.74	0.00	0.12	0.14
rub	0.00	0.76	0.11	0.12
scratch	0.01	0.15	0.79	0.05
tap	0.01	0.04	0.04	0.91

Figure 10. Confusion matrix for *action* classification results.



	big	bottomleft	bottomright	small	topleft	topright
big	0.46	0.01	0.02	0.32	0.13	0.06
bottomleft	0.02	0.41	0.12	0.01	0.21	0.24
bottomright	0.08	0.17	0.30	0.05	0.15	0.26
small	0.21	0.00	0.00	0.63	0.13	0.03
topleft	0.03	0.11	0.03	0.00	0.58	0.25
topright	0.04	0.11	0.10	0.03	0.23	0.50

Figure 11. Confusion matrix for *position* classification results.

interaction, and summed for diffusion on loudspeakers underneath the table.

The convolution and summing is performed by the `multconvolve~` object from the HISSTools Impulse Response Toolbox for modular creative IR manipulation [15, 16]. It implements efficient zero latency partitioned convolution in a matrix where each cell can have a separate IR.

6. ATTACK DETECTION

One drawback to convolution-based synthesis is the fact that the spectral contents of the piezo signal and the IR are multiplied. This means that spectrally poor (dull) interactions apply a low-pass filter to the IR. This is a limitation for using bright sounds also for subtle interactions. In order to overcome this and to add more sparkling high-frequency content, we decided to also detect attacks using the `bonk~` FFT-based attack detector for Max/MSP. A detected attack (that can come even from dull hitting) simply sends a dirac impulse with amplitude proportional to the volume of the attack through the convolver, effectively triggering the IR like a sample.

This creates another layer of interaction that is quickly discovered: depending on the contact interaction, one can summon a certain sound that is then available for triggering by hitting. It also opens up the possibility to add surprise sounds that are triggerable only part of the time (depending on the narrative during a visitors' session) or when certain external conditions are met.

7. CONCLUSIONS AND FUTURE WORK

Although the results of classification are not perfect, the musical interaction with the tables always works, since the expressive articulation of the composed sounds always takes place, even when sometimes a different sound than intended is used as IR. Together with the triggering of IRs

Figure 12. The finished table with chairs.

or other samples, the interaction is rich, and invites to discover all sorts of gestures and combinations of objects.

The next steps will be to work on the feature analysis of the piezo signal: discrimination between actions could be improved by also considering the temporal aspects of the signal, possibly by using the deltas of the MFCC frames.

The interactions that were trained by the GMM in the final installation [7] were only a subset of the test database, namely: rubbing on the centre pane with a plate, scratching outside with cutlery, tapping anywhere with cutlery. However, training was done with separate recordings for each of the 6 places around a table to account for the different shapes and texturing of the surfaces. This opens up a possibility of training a more discriminative classifier specially on these useful combinations, in order to increase accuracy and robustness.

Acknowledgments

This work has received support from the RAPID-MIX project (H2020-ICT-2014-1 Project ID 644862), funded by the European Union's Horizon 2020 research and innovation programme. We'd like to thank the team organising and assisting with the exhibit, Odile Fayet, Julie Gravier, Felicité Wintenberger, Brendan MacFarlane, Axel Adam, Taehoon Park, Dominique Jakob, Patrice Richard, Fabien Lepage, Emmanuel Flety, Arnaud Richer, Julien Pittet, Jérémie Henrot.

Figure 13. The finished table in use.

8. REFERENCES

[1] M. Savary, D. Schwarz, and D. Pellerin, "DIRTI — Dirty Tangible Interfaces," in *Proceedings of the Conference for New Interfaces for Musical Expression (NIME)*, Ann Arbor, MI, USA, May 2012, pp. 347–350.

[2] M. Savary, D. Schwarz, D. Pellerin, F. Massin, C. Jacquemin, and R. Cahen, "Dirty Tangible Interfaces: Expressive Control of Computers with True Grit," in *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI'13. New York, NY, USA: ACM, 2013, pp. 2991–2994. [Online]. Available: http://doi.acm.org.gate6.inist.fr/10.1145/2468356.2479592

[3] F. Bevilacqua, N. Schnell, N. Rasamimanana, B. Zamborlin, and F. Guédy, "Online Gesture Analysis and Control of Audio Processing," in *Musical Robots and Interactive Multimodal Systems: Springer Tracts in Advanced Robotics Vol 74*, J. Solis and K. C. Ng, Eds. Springer Verlag, 2011, pp. 127–142. [Online]. Available: http://articles.ircam.fr/textes/Bevilacqua11b/

[4] N. Rasamimanana, F. Bevilacqua, N. Schnell, F. Guedy, E. F. Come Maestracci, B. Zamborlin, and J.-l. F. Uros Petrevsky, "Modular Musical Objects Towards Embodied Control Of Digital Music," in *Tangible Embedded and Embodied Interaction*, 2011. [Online]. Available: http://articles.ircam.fr/textes/Rasamimanana11a/

[5] B. Zamborlin, F. Bevilacqua, M. Gillies, and M. d'Inverno, "Fluid gesture interaction design: applications of continuous recognition for the design of modern gestural interfaces," *ACM Transactions on Interactive Intelligent Systems*, vol. 4–3, pp. 30–45, 2014.

[6] B. Zamborlin, "Studies on customisation-driven digital music instruments." Ph.D. dissertation, Goldsmiths College, University of London, UK and Université Pierre et Marie Curie, Paris France, 2015.

[7] D. Schwarz, P. A. Tremblay, and A. Harker, "Rich Contacts: Corpus-Based Convolution of Audio Contact Gestures for Enhanced Musical Expression," in

---

[7] See figures 12 and 13 and the short video at http://www.dailymotion.com/video/x5s7wbd.

Proceedings of the International Conference on New Interfaces for Musical Expression 2014. Goldsmiths, University of London, UK: Goldsmiths, University of London, UK, 2014.

[8] D. Schwarz, "The Sound Space as Musical Instrument: Playing Corpus-Based Concatenative Synthesis," in *Proceedings of the Conference for New Interfaces for Musical Expression (NIME)*, Ann Arbor, MI, USA, May 2012, pp. 250–253.

[9] M. Puckette, "Infuriating Nonlinear Reverberator," in *Proceedings of the International Computer Music Conference (ICMC)*, Huddersfield, UK, 2011. [Online]. Available: http://hdl.handle.net/2027/spo.bbp2372.2011.090

[10] E. Tomás and M. Kaltenbrunner, "Tangible Scores: Shaping the Inherent Instrument Score." in *NIME*, 2014, pp. 609–614.

[11] E. Tomás, "Tangible Scores," in *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, 2016, pp. 669–674.

[12] N. Schnell, A. Röbel, D. Schwarz, G. Peeters, and R. Borghesi, "MuBu & Friends – Assembling Tools for Content Based Real-Time Interactive Audio Processing in Max/MSP," in *Proceedings of the International Computer Music Conference (ICMC)*, Montreal, Canada, Aug. 2009.

[13] N. Schnell, D. Schwarz, and J. Larralde, "PiPo, A Plugin Interface for Afferent Data Stream Processing Modules," in *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, Suzou, PRC, Oct. 2017.

[14] J. Françoise, N. Schnell, R. Borghesi, and F. Bevilacqua, "Probabilistic Models for Designing Motion and Sound Relationships," in *Proceedings of the 2014 International Conference on New Interfaces for Musical Expression*, London, UK, United Kingdom, Jun. 2014, pp. 287–292. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01061335

[15] A. Harker and P. A. Tremblay, "The HISSTools Impulse Response Toolbox: Convolution for the Masses," in *ICMC 2012: Non-cochlear Sound*, M. Marolt, M. Kaltenbrunner, and M. Ciglar, Eds. The International Computer Music Association, July 2012, pp. 148–155. [Online]. Available: http://eprints.hud.ac.uk/14897/

[16] ——, "Rethinking the Box: Approaches to the Reality of Electronic Music Performance," in *IRCAM Forum*, November 2013. [Online]. Available: http://eprints.hud.ac.uk/18549/