The Talking Guitar: Headstock Tracking and Mapping Strategies

Liam B. Donovan
Media and Arts Technology, School of EECS
Queen Mary University of London
Mile End Road, London E1 4FZ, UK
l.b.donovan@qmul.ac.uk

Andrew P. McPherson
Centre for Digital Music, School of EECS
Queen Mary University of London
Mile End Road, London E1 4FZ, UK
a.mcpherson@qmul.ac.uk

ABSTRACT

This paper presents the Talking Guitar[1], an electric guitar augmented with a system which tracks the position of the headstock in real time and uses that data to control the parameters of a formant-filtering effect which impresses upon the guitar sound a sense of speech. A user study is conducted with the device to establish an indication of the practicality of using headstock tracking to control effect parameters and to suggest natural and useful mapping strategies. Individual movements and gestures are evaluated in order to guide further development of the system.

Keywords

Augmented guitar, formant filtering, headstock tracking.

1. INTRODUCTION

Guitar effects have existed for nearly as long as the electric guitar itself, but until recently, performers had few options for controlling their parameters while playing. Switches and knobs are difficult to operate during performance, and expression pedals offer only a single dimension of control. In recent years, interest has grown in *augmented guitars* [10] which incorporate new controls onto the instrument itself. Many augmentation approaches are possible, but in all cases, it is important to consider how the new controls relate to existing performance technique.

This paper presents a study of guitar augmentation using headstock tracking. An illuminated sphere attached to the headstock (Figure 1) is tracked by camera, and the resulting position is used to control a vocal formant filtering effect. We present an analysis of performer interaction with the instrument, focusing on the role that the choice of mapping plays in the performer's patterns of motion.

2. BACKGROUND

2.1 Augmented Guitars

Because of its hybrid electric/acoustic nature the electric guitar is well suited to being augmented with technology to increase musical options for the player. Listed here are several relevant recent examples.

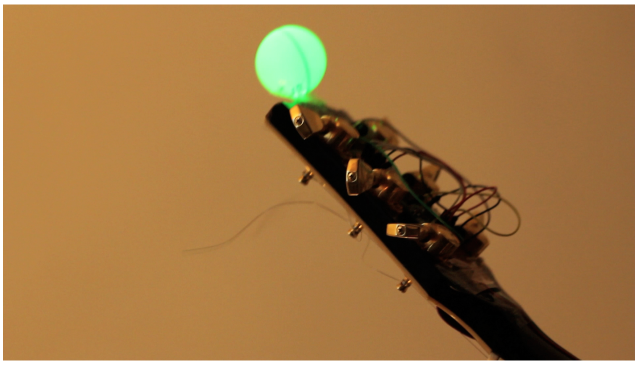

Figure 1: An illuminated sphere on the headstock tracked by camera to control a formant filter effect.

UniComp [6] uses a smartphone strapped to the wrist of the performer to provide several methods of wirelessly controlling effect parameters including on-screen sliders and the built-in accelerometer. The RANGE guitar [13] is a self-contained system which uses touch based linear potentiometers mounted on the body of the guitar to control parameters of effects which are implemented with an on-board embedded micro-processor. The Augmentalist [14] is a system which allows musicians to augment their own guitar with sensors, and develop their own mappings for those sensors. Observed configurations included mapping neck tilt angle to effect parameters with an accelerometer. The most prominent augmented guitar in the literature is the multimodal guitar [10, 15], which uses audio features extracted from the guitar signal along with inputs from several sensors measuring the tilt angle of the neck and pressure on the bridge, a touch-sensitive linear potentiometer and several contact microphones on the guitar to affect the sound.

2.2 Formant Filtering and Vocal Effects

One of the most popular and widely used guitar effects is the wah pedal, so called because it uses a single resonating filter which can be swept up and down the frequency spectrum with a foot pedal creating a vocal-like "wah" sound. Attempts to produce new control interfaces for the wah effect have used the player's voice [12], and the effect has been extended to have multiple peaks with static relationships as in the M-fold wah [3], but despite its popularity the effect is limited to only one vowel sound and one dimension of control. The aim of the device presented here is to give the guitar player more dimensions of control, more vowel sounds to choose from and an intuitive way to choose them, allowing the expressivity of the effect to be increased.

Humans recognise vowels by concentrations of energy around certain frequencies in the spectrum of a vocalised sound which are independent of the owner or the pitch of the voice. These concentrations are a product of the reso-

---

[1] Video: `http://vimeo.com/57209669`
`http://www.liamdonovan.co.uk/#talkingguitar`

NIME'14, June 30 – July 03, 2014, Goldsmiths, University of London, UK.

nant filtering effect the vocal tract has on the sounds produced by the vocal chords [4], and their central frequencies are called formants. In real speech between five and six formants are produced, however only the lowest three are necessary to produce vowel sounds [9].

There are several examples of performance orientated musical speech synthesisers in the literature including FutureGrab [5] which generates vowel sounds through formant filtering controlled by a set of sensor-filled gloves, mapping hand shapes to vowel sounds; and HandSketch [2], an augmented graphics tablet for synthesising expressive vocals.

In a recent paper [1], Astrinaki et al. fused two separate strands of research to use an augmented guitar as a controller for a speech synthesis system called MAGE by extracting audio features from the guitar signal.

2.3 Musical Gestures

Musical gestures can be categorised under the following four typologies: Sound producing, sound accompanying, sound facilitating and communicative [8]. Lähdeoja et al. have done some excellent work on harnessing sound-facilitating gestures in electric guitar playing (also classified as ancillary gestures and defined as "gestures that are not directly involved in sound production, but still play an important part in shaping the resultant sound" [8]); using them to control guitar effect parameters in a subtle and indirect way without conscious effort by the player [11].

3. SYSTEM DESCRIPTION

3.1 Headstock Tracking

In this paper we have created a system which uses guitar headstock tracking to control a formant filtering effect. The tracking is achieved by attaching a green LED to the guitar inside a ping-pong ball to diffuse the emitted light, and using Jitter to track the green blob across the 2D plane of a webcam feed. This solution was found to be suitably robust with low enough latency for the purposes of this study.

3.2 Formant Filtering

The filter stage works by manipulating the centre frequencies and bandwidths of three cascaded peaking-filters to emulate the vocal tract and produce formants, impressing recognisable vowels onto the guitar sound. Nine vowel sounds and four sonorant consonants (those which can be simulated in a similar way to vowels with three formant filters) are used in the effect to give a variety of speech-like sounds. The consonants used are /w/, /y/, /r/ and /l/; the vowel sounds, selected from those defined in [9], are /a/, /iy/, /ey/, /ae/, /ow/, /uw/, /ay/, /aw/, and /oy/.

3.3 Mapping

For this system we used a direct mapping between headstock position and filter response. The strategy was to have the central position where the user holds the guitar naturally correspond to one of the four sonorant consonants, and then to have the vowels arrayed on the radius of a circle centred on the consonant. Thus moving the guitar from the centre of the screen to the edge causes the filters to change from the consonant sound to the vowel sound, and circling the guitar around the screen moves between vowel sounds (see Figure 2).

The mapping system, implemented in Max/MSP, generates the formant frequencies and bandwidths by first interpolating between the two nearest vowel sounds, and then interpolating between those values and those of the central sonorant depending on the distance from the headstock to the centre of the screen. This allows a smooth variation between the different sounds.

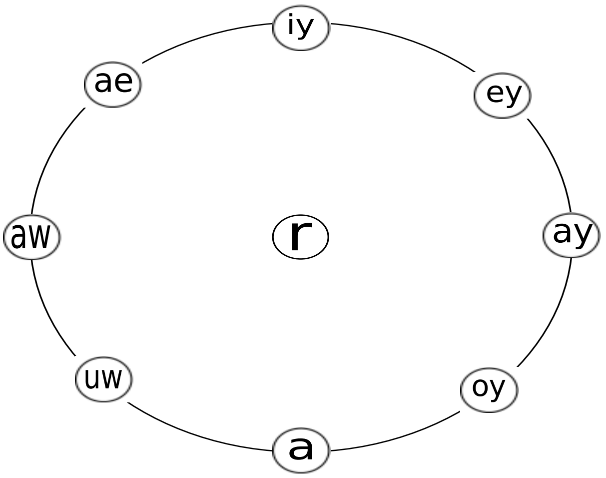

Figure 2: A diagram showing the 8-vowel mapping. All other mappings were subsets of this.

4. DISCUSSION

4.1 Mapping

It has long been known that without a natural, intuitive and expressive mapping between physical and audio parameters, an electronic instrument can not pretend to be musically useful [7]. With this device, the input is an augmentation to traditional guitar technique, and so the mapping between the headstock position and the effect parameters must satisfy the demands laid out above whilst not distracting the player from either the physical or mental challenge of playing the guitar. If the effect takes too much cognitive bandwidth, the performance will suffer.

In the user study presented in this paper we provided participants with five different mapping configurations. For all the mappings, the centre of the screen was mapped to the /r/ sonorant. For the first two mappings two vowels were first arranged horizontally (2h) and then vertically (2v); for these only movement in the x or y directions caused any change in the effect, movement perpendicular to these directions did nothing. For the following two mappings four vowels were arranged in a + configuration (4+) and then a X configuration (4x), and for the final mapping eight vowels were distributed equally around the circumference of a circle (8) (see Figure 2). For these last three configurations, the system used the polar interpolation method described above in (3.3).

4.2 Non-Intrusive Gestures

In order not to interfere with normal playing technique the ideal gesture typology to use for deliberate control of effect parameters is one as far decoupled from sound production mechanisms as possible. It has been suggested that large upper-body movements of guitarists and of the guitar neck are largely disconnected from sound production [15], and it seems intuitively true as guitar-playing technique is dependant upon precise movement of the fingers rather than large movements of the hand, arm and neck, as can be seen by the use of the neck as a communicative gesture by many performing guitarists. The use of such a communicative gesture is interesting because it is already used by guitarists to connect the audience to the sound in performance. Using the gesture to control a new interface may help to avoid a common pitfall of many new electronic instruments - that is the lack of perception or understanding from the audience of the connection between a performers' actions and the resulting sound [16].

4.3 Study Goals

The goal of this study is to establish an intuitive means of control for the effect described above. The first objective is to determine whether gestures involving moving the headstock of the guitar are sufficiently decoupled from sound production mechanisms for it to be possible to use them to deliberately control the effect without obstructing the playing of the guitar. Further objectives include establishing the usefulness of more specific gestures which come under the umbrella of 'headstock movement', and establishing the most intuitive mappings between position and vowel sounds for this effect.

5. USER STUDY

The six participants were first given a short time to familiarise themselves with the guitar to be used in the experiment. Movement was tracked with a webcam oriented to the left of the participants such that when they were standing comfortably, the guitar neck was pointing directly towards the camera and the headstock was directly in the centre of the camera's field of view. With this orientation the movements of the guitarists were projected onto the plane of the camera in such a way that when the guitarist tilted the neck up, the ping pong ball moved up on the screen, and when the participant walked forward or twisted their body or the guitar clockwise, the ping-pong ball moved right on the screen.

The participants were first asked to play for a minute without the effect engaged and were then asked to play for two minutes with each of the five different mappings described in (4.1). For the first of their two minutes with each effect, participants were asked to play in an exploratory way, moving slowly about the space and lingering on each of the vowel sounds. For the second minute, they were asked to play the effect with as much expression as possible. For all playing periods both the audio and the movements of the headstock were recorded. An example plot of one participant's session with several mappings can be seen in Figure 4. After the experiment, the participants were presented with a short survey asking them to classify the difficulty, range and expressivity of each mapping and name their favourite.

6. RESULTS

6.1 Headstock Gesture Analysis

The headstock movement data was analysed by taking the mean distance from the mean position of the headstock both when the effect was in use and not. Of a maximum distance of 0.5, the mean distance with the effect was 0.22 and without it was 0.03. This much larger movement with the effect engaged suggests that large headstock-movement gestures are neither necessarily sound-producing nor ancillary gestures, and therefore have the potential to be available for use as control inputs for the effect.

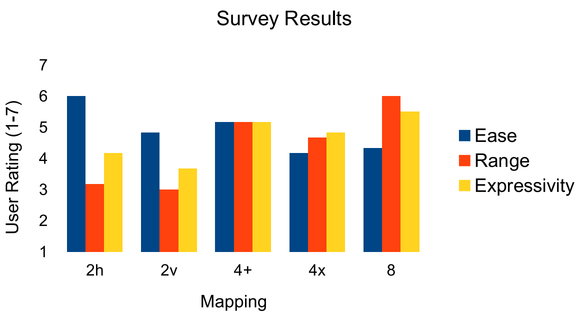

Figure 3: Results from the user survey.

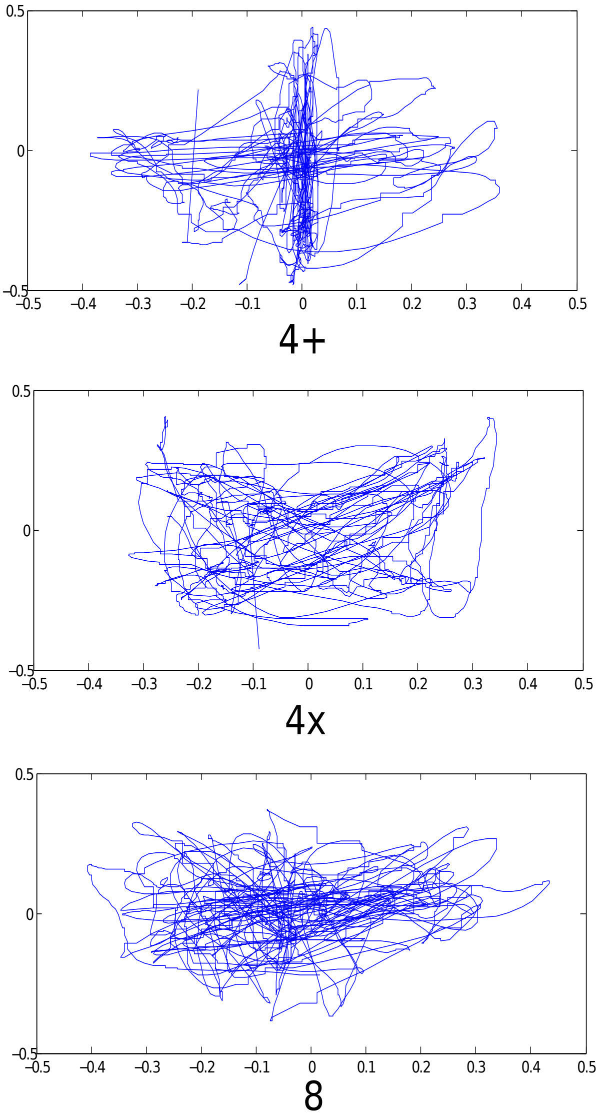

Figure 4: A plot of one participant's headstock movements using the three polar mappings.

6.2 Survey Analysis

The participants were asked to rank each of the mappings on how easy they felt it was to play, what range of sounds they felt they were able to controllably elicit from the effect, and how expressive they felt the mapping was; the results are plotted in Figure 3. It can be seen that the general trend is towards more complex mappings with more vowels being more difficult to play but more expressive and offering a greater range of sounds to the user. This is backed up by the fact that three participants chose the 8-vowel mapping as their favourite, compared with two for the 4+ mapping and one for the 4x mapping.

A possible explanation for this trend is that the guitar neck is heavy, precision in the headstock gestures is tricky, and any large movement carries significant physical and cognitive cost. A mapping which offers a larger variation in sound for a smaller movement therefore feels more expressive, even if that is at the cost of precision, which for this effect is perhaps not essential.

Another observation from the survey is that the three polar mappings outweigh the linear mappings in terms of expressivity and range, but that of those the off-axis mapping (4X) is the most difficult and least expressive.

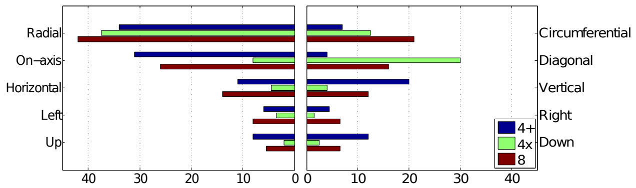

Figure 5: Mean number of sonorant transitions across all participants for the three polar mappings.

During the survey, three of the participants stated they felt it was difficult to play the guitar and control the effect simultaneously - of those two suggested this ability would improve with practice, with the other feeling it may just be too hard to do both.

6.3 Sonorant Transition Analysis

For each mapping the headstock position data was analysed to count the frequency of the transitions between the regions corresponding to the sonorants. The results of this analysis are displayed in Figure 5; a radial transition is defined as one from the central consonant to a vowel around the perimeter, a circumferential transition is defined as a movement from vowel to vowel around the perimeter, an on-axis transition is defined as a radial transition along the axes, and a diagonal transition is defined as a radial transition along the diagonals.

The results show a clear preference throughout the experiment for radial over circumferential transitions and for on-axis over diagonal transitions, which suggests that radial, on-axis movements are the easiest and most intuitive. This is backed up by the survey data, in which the 4+ mapping is shown to be easier and more expressive than the 4x mapping, and several comments in the survey suggesting that on-axis movements were more closely aligned to the participant's natural movements. This was not universal however, one participant expressly preferred the 4x mapping, and felt the diagonals were more aligned to their natural movements.

Another interesting observation is that the ratio of circumferential to radial transitions significantly increases between the four-vowel mappings and the eight-vowel mapping. In the eight-vowel mapping the distance around the circumference between the vowels is half as big as in the four-vowel mappings, and the circumferential distance between the vowels is smaller than the radial distance between the central consonant and the vowels. This seems to act as an incentive to produce circumferential movements, as the ratio between them and radial movements is roughly twice as big in the eight-vowel mapping.

This supports the notion that because the guitar neck is heavy and the cost of movement is high, the participants preferred large variations in the effect for small movements of the headstock, even if those variations required the more difficult circumferential gesture, and even if they came at the cost of precise control.

7. CONCLUSIONS

We have presented a user study of the Talking Guitar, an electric guitar augmented with headstock tracking to control a formant-filtering effect, contrasting several different mapping configurations. We analysed the results for indications of whether headstock tracking is a useful or distracting augmentation, and looked at which specific movements and mappings are most useful. It was suggested that headstock-moving gestures are not directly coupled to the production of sound in guitar playing, and so do have the potential to be useful as control gestures for effect parameters. It was noted however that they do carry a significant physical and cognitive cost. It was suggested that on-axis radial gestures may carry a lower cost than diagonal or circumferential ones, and that more complex, densely-packed mappings were preferred as more expressive. The results were interpreted to suggest that due to its high physical and cognitive cost, the headstock-moving gesture is more suited to controlling large variations in effect parameters with smaller movements, as opposed to precise variations with large movements.

8. ACKNOWLEDGEMENTS

This work was supported by the EPSRC Centre for Doctoral Training in Media and Arts Technology, Queen Mary University of London.

9. REFERENCES

[1] M. Astrinaki, L. Reboursière, A. Moinet, N. D'alessandro, T. Dutoit, et al. Is this guitar talking or what!? In *Proc. eNTERFACE'12*, 2012.

[2] N. d'Alessandro and T. Dutoit. Handsketch bi-manual controller: Investigation on expressive control issues of an augmented tablet. In *Proc. NIME'07*, pages 78–81. ACM, 2007.

[3] S. Disch. Digital audio effects-modulation and delay lines. *Master's thesis, Technical University Hamburg-Harburg*, 1999.

[4] G. Fant. *Acoustic theory of speech production.* Number 2. Walter de Gruyter, 1970.

[5] Y. Han, J. Na, and K. Lee. Futuregrab: A wearable synthesizer using vowel formants. *NIME'12*, 2012.

[6] O. Hödl and G. Fitzpatrick. Exploring the design space of hand-controlled guitar effects for live music. In *Proc. ICMC*, 2013.

[7] A. W. Hunt and M. Paradis. M.(2002)"the importance of parameter mapping in electronic instrument design.". In *Proc. NIME'02.*

[8] A. R. Jensenius, M. M. Wanderley, R. I. Godøy, and M. Leman. Musical gestures. *Musical Gestures: Sound, Movement, and Meaning*, page 12, 2009.

[9] D. H. Klatt. Software for a cascade/parallel formant synthesizer. *the Journal of the Acoustical Society of America*, 67:971, 1980.

[10] O. Lähdeoja. An approach to instrument augmentation: the electric guitar. In *Proc. NIME'08*, page 53, 2008.

[11] O. Lähdeoja, M. M. Wanderley, and J. Malloch. Instrument augmentation using ancillary gestures for subtle sonic effects. *Proc. SMC*, pages 327–330, 2009.

[12] A. Loscos and T. Aussenac. The wahwactor: a voice controlled wah-wah pedal. In *Proc. NIME'05*, pages 172–175. National University of Singapore, 2005.

[13] D. MacConnell, S. Trail, G. Tzanetakis, P. Driessen, W. Page, and N. Wellington. Reconfigurable autonomous novel guitar effects (range). In *Proc. SMC*, 2013.

[14] D. Newton and M. T. Marshall. The augmentalist: Enabling musicians to develop augmented musical instruments. In *Proc. TEI'11*, pages 249–252. ACM, 2011.

[15] L. Reboursière, C. Frisson, O. Lähdeoja, J. Anderson, M. Iii, C. Picard, and T. Todoroff. Multimodal guitar: A toolbox for augmented guitar performances. In *Proc. NIME'10*, 2010.

[16] J. Schacher. Action and perception in interactive sound installations: An ecological approach. In *Proc. NIME'09*, 2008.