

LNAI 9183

Dylan D. Schmorrow
Cali M. Fidopiastis (Eds.)

Foundations of Augmented Cognition

9th International Conference, AC 2015
Held as Part of HCI International 2015
Los Angeles, CA, USA, August 2–7, 2015, Proceedings



 Springer

Augmenting Bioacoustic Cognition with Tangible User Interfaces

Isak Herman¹(✉), Leonardo Impett², Patrick K. A. Wollner²,
and Alan F. Blackwell¹

¹ Computer Laboratory, University of Cambridge, 15 J. J. Thompson Ave.,
Cambridge CB3 0FD, UK

`isak.herman@cl.cam.ac.uk`

² Department of Engineering, University of Cambridge, Cambridge, UK

Abstract. Using a novel visualization and control interface – the Mephistophone – we explore the development of a user interface for acoustic visualization and analysis of bird calls. Our intention is to utilize embodied computation as an aid to acoustic cognition. The Mephistophone demonstrates ‘mixed initiative’ design, where humans and systems collaborate toward creative and purposeful goals. The interaction modes of our prototype allow the dextral manipulation of abstract acoustic structure. Combining information visualization, timbre-space exploration, collaborative filtering, feature learning, and human inference tasks, we examine the haptic and visual affordances of a 2.5D tangible user interface (TUI). We explore novel representations in the audial representation-space and how a transition from spectral to timbral visualization can enhance user cognition.

Keywords: Tangible user interfaces · Embodied interaction · Bioacoustics · Information visualization · Collaborative filtering

1 Introduction

Initially devised for a case-study exploring the dynamics of modern music production and performance, we have now designed and built a generalized device capable of a novel set of modes for human and computer interaction with sound, space, and light. The Mephistophone is a haptic interaction device combining sensors for surface-depth data and local illumination levels along with controlling actuators which deform the shape of its latex surface. The computational architecture supports real-time algorithms capable of generating novel and familiar mappings of movement and light from sound features and vice versa. In this paper, we briefly outline the design concept and introduce a specific application of this novel device as a mechanism for enhancing user cognition of bird calls through high dimensionality information visualization. We have described the low-cost design and development of the initial prototype and specified the technical challenges in creating a novel, large-scale interaction controller from scratch [24]. We have classified potential interactions with the tangible user

interface (TUI) as the triggering, controlling, editing, and manipulation of sounds with varying temporal scope and structure. The system allows the incorporation of collaboratively filtered end-user gesture behavior performed in reaction to sound samples to inform the behavior of the physical and visual representation presented to subsequent users. The design is premised on the concept of a TUI which satisfies the metaphor of instrument-as-participant, rather than instrument-as-tool. As a user interacts with the device, it becomes aware of correlates between perception of sound and shape; with time the Mephistophone comes to augment the users intentions, guiding them towards learned pathways reflecting prior users' interactions.

We are exploring the future development of haptic TUIs and embodied computation. Our goal with the Mephistophone is to question the dynamics of acoustic visualization and analysis through a technically advanced prototype that demonstrates "mixed initiative" design, where humans and algorithms together collaborate toward creative and purposeful goals and where the interaction modes allow direct manipulation of abstract structure [5]. For this case-study, we selected the bioacoustic domain for bird calls express high timbral variation which is not captured using current information visualization techniques. Using a set of training gestures from prior research that were generated in response to timbre based samples, we explore physical mappings from extracted acoustic features and have augmented the system with new dimensions of visual feedback. In ongoing research we are comparing the efficacy of standard acoustic features with learned feature sets extracted from environmental audio for user cognition in the physical and visual domains.

2 Prior Work

Our investigation arrives at a confluence of research paths which treat the visual and auditory domains in analogous fashion. Through this research, we explore how connections made by both the system and the user from one domain or with one sense may inform interpretation of the other demonstrating mixed-initiative design.

2.1 Timbre Spaces

Timbre refers broadly to those features of sound which cannot be summarized by frequency and amplitude. With such an opaque definition, timbre does not easily lend itself to visual representation. In the iterative design of the device, we realized that the affordances of the haptic tangible interface, whilst novel as an interaction mechanism, provided similar dimensionality reduction requirements as with information visualized on a conventional 2D color display. This led to the current design which can overlay an equivalent 2D visual interface on our physical surface decreasing the need for such dimensionality reduction.

The majority of bioacoustic research into the analysis of audio data has been performed using temporal and spectral information, visualized as spectrograms. The time/frequency resolution trade-offs implicit to such models, as well

as the limitations of augmenting user cognition of audio signals given only time-series frequency and amplitude data, left us considering the role of timbre in audio cognition and means by which temporally relevant timbral feature patterns could be visually depicted simultaneously with familiar spectral features. In attempting to augment novice users' bioacoustic cognition based on visual representation of timbral information, we explore means to offer 'implicit learning processes [which] enable the acquisition of highly complex information in an incidental manner and without complete verbaliz[able] knowledge of what has been learned' [23].

Timbre remains the audio component least quantified across a variety of research approaches; in various works it has been proposed to comprise a combination of attack time, spectral centroid, spectral flux, and spectrum fine structures [1]. It is this information which we endeavor to visualize as an augmentation to spectral representations of audio data. Researchers from Institut de Recherche et Coordination Acoustique/Musique (IRCAM) have proposed various means to visualize 'the structure of the multidimensional perceptual representation of timbre (the so-called timbre-space) . . . and then attempt to define the acoustic and psychoacoustic factors that underlie this representation' [11]. Early psychoacoustic modeling work pursued the use of multidimensional scaling (MDS) to obtain a timbre-space from which acoustic correlates of the dimensions could be identified and predictive models built. They state that 'timbre spaces represent the organization of perceptual distances, as measured with dissimilarity ratings, among tones equated for pitch, loudness and perceived duration' along with spatial location and the environmental reverberance [1]. They further resolved the statistical problems arising from the fact that such models were defined so that each listener added to their data set increased the number of parameters in the model by proposing that listeners' perceptions of timbre would fall into a pre-defined, comparatively small, number of latent classes [11]. These models have shown promise, insofar as they offer replicable dissimilarity judgments between classes. Experiments based on finding associations between these acoustic correlates of timbre (flux, centroid and attack) and proposed neural correlates of timbre have seen support in fMRI studies showing greater activation in response to increases in flux and centroid combined with a decreased attack time [12]. For our novel visualization interface, we considered not only the need to augment the traditional spectral representations of audio with timbral information, but the likelihood that users, whilst not initially cognoscente of our timbral visualization mapping, would through exploration be able to build a mental model between the visual representation of our timbral model and a set of latent timbral class representations which could show correlation to the timbral attributes of attack time, spectral centroid, and spectrum fine structure.

2.2 Information Visualization

The instrument-as-participant characterization of the Mephistophone from its earliest design encouraged us to explore mechanisms by which information visualized, as well as informed as gestures, on its surface could map to known acoustic

features. With the augmentation of illumination we maintain the capacity for replicating familiar acoustic information visualization metaphors whilst retaining additional degrees of freedom. Recent frameworks for information visualization theory have proposed that ‘a model of data embodied in a visualization must be explored, manipulated and adapted ... [and w]hen there are two processing agents in a human-computer interaction context ... either or both can perform this processing.’ We research the ‘interpretation of a visualization through its physical form’ as collaboratively filtered by the user, as well as the ‘exploration and manipulation of the internal data model by the system in order to discover interrelationships, trends and patterns’ [17] in our acoustic datasets. Purchase et al. further note that ‘a prediction that a dataset may contain groups (clusters) of objects with similar characteristics does not define what specific clusters there are ... [but] orients tool designers who will know that the tool must help the users to detect clusters.’ Although we consider the Mephistophone a participant in the interaction rather than a tool, these ideas guide our design in selecting a space of exploration in which latent timbral and gestural classes are modeled to arise. Early representations of acoustic data on 2D interfaces began mitigating the limitations of such displays to convey high dimensionality feature data by taking advantage of multi-dimensional color features such as the RGB *tristimulus equivalent* forming a *chromaticity vector* as augmentations to the physical axes of the display [9]. Spectrograms, which have formed the primary mode of visualization for the features of an acoustic signal both in bioacoustic informatics and other audio domains, are limited to the spectral to the detriment of conveying timbral information.

Acoustic Visualization. Visualizations of acoustic data have predominantly been used to describe spectral information over time, from early grayscale implementations, to those using the RGB tristimulus dimensions of color as previously noted. With the Mephistophone, we have the capacity to retain all such spectrotemporal features and convey further information on the unused tangible dimensions or map those features to the tangible interface retaining the illumination space for a novel mapping. Within bioacoustics, be it research into general biodiversity metrics, or single species identification, spectral representations remain the most investigated. This is a result of researchers’ familiarity with the spectrogram visual representation rather than an implicit result of such representations being superior. Recent research into the efficacy of temporally brief sound event (as opposed to speech) classification, which better reflects our target audio when working with bird calls, has identified ‘distinctive time-frequency representations ... motivated by the visual perception of the spectrogram image’ which work with grayscale spectrograms to reduce noise [3]. While noise reduction remains important, such methods eliminate significant information contained within the original audio signal.

We propose that methods of timbral visualization, used previously in musical analysis to supplant rather than augment standard spectral visualizations, offer valuable additional information of use to a novice user or automated recognition

system. The additional dimensions available with the Mephistophone, allow us to map timbral information as well as spectral onto the same interface. Early work from the Center for Computer Research in Music and Acoustics (CCRMA) at Stanford University has proposed mapping the physical parameters of timbre to a constrained feature set derived from Fourier output including power, bandwidth, centroid, harmonicity, skew, roll-off and flux [19]. Subsequent work by those researchers found support for a visual representation of a tristimulus based timbre descriptor “based on a division of the frequency spectrum in three bands [which] like its homonymous model for color description . . . provides an approximation to a perceptual value through parametric control of physical measurements” overcoming the time-frequency trade-off inherent to spectrogram visualizations [20]. Their multi-timescale aggregation ‘maps features to objects with different shapes, textures and color in a 2D or 3D virtual space’ [19]. We have taken this mapping from a virtual to tangible interface, concurrently maintaining the multi-scale temporal representations of their visualization, with additional dimensions available to visualize conventional time-series information. Several alternative multi-timescale aggregation visualizations have been developed which animate characteristics of timbral space including but not limited to centroid width and spread, skewness, flux, bark-flux, and centroid-flux, as well as roll-off [21]. As we are less constrained in dimensionality reduction when working with projection and animation on a 2.5D tangible interface, the Mephistophone allows us to overcome some of the temporal and dimensionality reduction limitations of a standard screen when simultaneously visualizing spectral and timbral information.

2.3 Tangible User Interfaces

Haptic, tangible interfaces have been devised for a variety of purposes, from physical modeling of landscapes to musical collaboration environments. The Mephistophone was designed to simultaneously display a representation of audio and allow the user to haptically interact with the surface for tangible real-time audio manipulation. Various projects have designed reconfigurable, haptic, input devices with the capacity to sense objects placed within, on, or near a device, using techniques ranging from capacitive sensing to optical approaches [7]. Our 2.5D display technology incorporates a number of features of previous work, albeit at significantly reduced cost, but, given our sensing and display techniques, without the issue of occlusion of information on the surface by the user.

Early work from the MIT Media Lab explored the ‘benefits of [a] system combining the tangible immediacy of physical models, with the dynamic capabilities of computation simulations’ where the user alters a model’s physical topography in a feedback loop with a computational system capable of assessing the changes brought about by the interaction [16]. Prior research in the domain of acoustic interaction explored the use of projection from beneath onto a surface, thus avoiding the issue of occlusion, whilst providing dynamic visual feedback to the users regarding the state of the device. As an extension of this

model with our current design, we combine projection with embedded illumination on the physical surface [8]. Another project informing our system ‘is both able to render shapes and sense user input through a malleable surface’ although they achieved actuation using electric slide potentiometers requiring significantly more embedded computational power than our construction for roughly equivalent rates of surface animation [10]. Most recently, another group from the MIT Media Lab constructed a shape display which ‘provide[s] multiple affordances over traditional displays namely graspable deformable screens and tangible output ... giving [the] functionality of linear actuator displays with the freeform interaction afforded by elastic surfaces to create a hybrid 2.5D display’ [2]. As an extension in their work, “a grammar of gestures has been implemented illustrating different data manipulation techniques” [2]. We propose instead a model where gesture classes are learned by the device through collaboratively filtered user interactions. We are exploring extensions to the acoustic spaces we can represent compared to those on a 2D display, such as traditional spectrograms, given the affordances of haptic over visual representations as well as intuitive 3D analysis with physical models.

2.4 Bioacoustics

Bioacoustics, research into the acoustic emissions of animals, offers an accessible method for performing semi-automated biodiversity monitoring offering ‘reproducible identification and documentation of species’ occurrences’ [14]. Rapid analysis of extensive recordings requires efficient methods of target identification and noise reduction. We propose that the simultaneous spectral and timbral display capabilities of the Mephistophone diminishes the expertise necessary for novice users to isolate regions of interest in a visual representation of sound when combined without our haptic filtering model.

Effective bioacoustic monitoring, combining autonomous static or mobile user-centric sensor networks for data collection with interfaces whereby novice users can participate in data assessment, requires less user training than current visual monitoring techniques although a role for experts in validation remains. In our research we consider avian species which are frequently acoustically conspicuous whilst visibly camouflaged. However, ‘comparatively high song variability within and between individuals makes species identification challenging for observers and even more so for automated systems’ [14]. This has guided us to the following approach:

1. As avian calls are less temporally diverse than songs, we focus on short, timbrally variable utterances. These are best suited to generating community rather than species specific information.
2. As automated systems benefit from the results of human inference, we propose a collaborative filtering model.
3. As validated user feedback in a collaborative learning framework is necessary, we propose the training phase incorporate expert, if avocational, users.

Acoustic biodiversity surveyors have proposed indices measuring acoustic entropy (α), correlated with number of species in a community, and temporal and spectral dissimilarity (β), correlated with variation across communities. ‘Total species diversity in a set of communities has been traditionally seen as the product of the average diversity within communities (α) and the diversity between communities (β)’ [22]. Extensions to this model have demonstrated that these indices yield results equivalent to visual approaches to biodiversity assessment with increased efficiency [4]. Avian research has explored ‘the suitability of the acoustic complexity index (ACI) ... an algorithm created to produce a direct quantification of the complex biotic songs by computing the variability of the intensities registered in audio-recordings, despite the presence of constant human-generated noise’ and note that:

‘The ACI formula is based on the assumption that biotic sounds ... have a great variability of intensity modulation, even in small fractions of time and in a single frequency bin [and propose that] this new methodology could be efficiently used in hi-fi soundscape investigation to provide an indirect and immediate measure of avian vocalization dynamics, both in time and space, even when conducted by observers with limited skills identifying bird song’ [6, 15].

Acoustic indices reinforce the validity of short time event analysis as a means of noise reduction but only propose clustering by spectral similarity. We propose that timbral similarity distance functions might be learned on a device such as the Mephistophone and provide a useful extension to classification from clusters.

3 The Mephistophone

The underlying mechanical structure and computational architecture of the Mephistophone have been described previously and outline our modular arrangement of actuators, sensors, and micro-controllers as well as an early model of information flow [24]. In initial experiments, audio features including pitch classes and mel-frequency cepstral coefficient (MFCCs) were time synced with physical measures of static gestures informed on the surface (see Fig. 1). A preliminary machine learning model applied principal component analysis (PCA) to the extracted audio feature space and inferred a mapping to the tangible gesture space. Due to the computational complexity of audio feature extraction in previous experiments exceeding the capacity of the on-board micro-controllers, the analysis has been performed off-line using prerecorded samples. Subsequently we have implemented a number of extensions to the 2.5D haptic tangible display previously described as well as refined multiple computational architectures specific to various tasks.

In the current prototype an efficient, albeit low resolution, Fourier analysis program provides real-time spectral feature extraction from microphone input and runs on one of the micro-controllers allowing the the Mephistophone to dynamically depict local soundscapes. For more complex interaction, a small

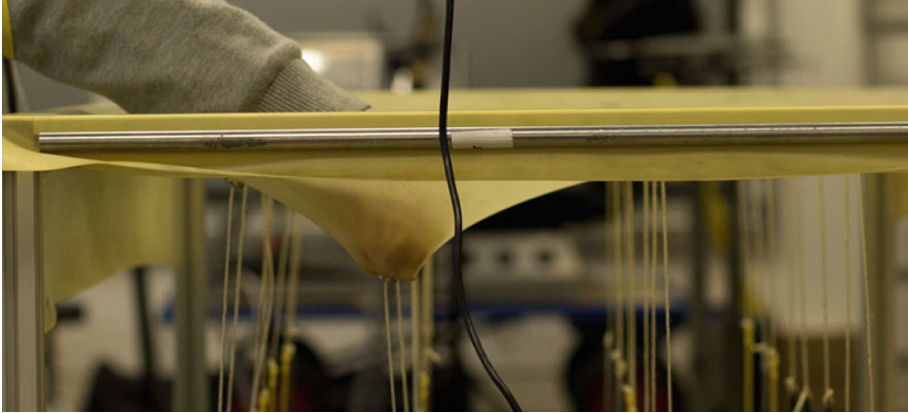


Fig. 1. Preliminary training where static gestures informed on the surface in response to audio cues are sensed.

computer mounted on the chassis provides augmented feature extraction capabilities. Incorporating further dimensions for feature representation on the device allows us to reduce the constraints on dimensionality reduction of sensed and analyzed features. In this iteration of the design, we have augmented the device with two mechanisms for displaying color allowing the mapping of more information from the acoustic feature space to the tangible and visual feature spaces. Base mounted projection capabilities allow color mapped representations to be displayed across the surface without occlusions, and surface mounted illuminated nodes allow fine-grained color-based representations of state to exist uniformly spaced across the surface (see Fig. 2). Upon incorporating both coarse and fine-grained illumination capabilities into the visual representation of the physical model, we retain the original 2.5D tangible space initially for depicting familiar features, and use the visual overlay on the surface to depict additional information. In our initial experiments, visual representation of analyzed audio features, including spectral bands and MFCCs, have been partitioned to be depicted on the device as physical and visual components; we treat illumination as a space for mapping timbral features overlaid on the physical representation of spectral features. In future work, we are developing a model whereby learned features are extracted from raw audio and mapped across both the physical and visual dimensions of the Mephistophone.

We use the Mephistophone to examine the potential of a novel 2.5D haptic representation of avian vocalizations amidst environmental soundscapes as a tool for bioacoustic analysis and biodiversity assessment; additionally we explore motivating novice user interaction with representations of various feature sets. The mechanical characteristics of the device allow more features extracted from the raw audio to be mapped than with a standard 2D color display extending bioacousticians current data visualization frameworks. As a visualizer, the Mephistophone can depict temporal patterns in the timbre-space of a soundscape,

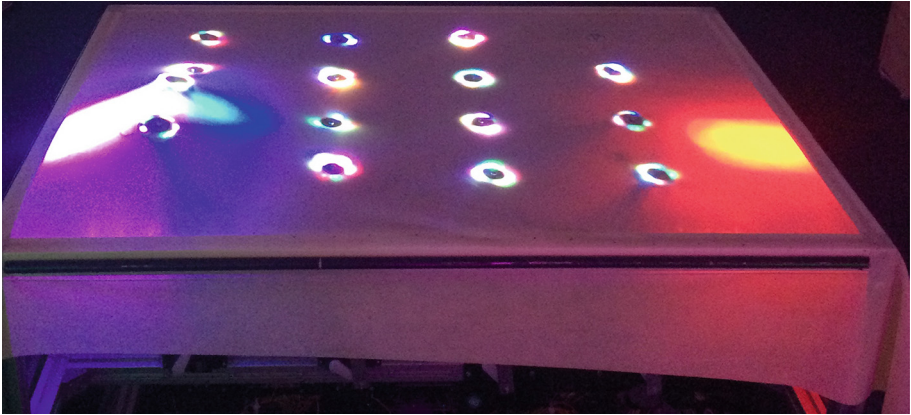


Fig. 2. Coarse and fine-grained illumination of the tangible interface

subsuming, supplanting, or combining with spectral patterns to provide multi-timescale representations unavailable with traditional displays. As we augmented the design with illumination and projection capabilities designed to reactively change the visual information presented on the physical surface, we augment standard bioacoustic visualizations and offer more simultaneous output dimensions. By removing the user from the familiar metaphor of 2D spectral representations of sound, the surface contains information about the original signal not available in traditional visualizations and elicits motivating curiosity in avocational users. With this design, dimensionality reduction algorithms need not be as constrained in output as would normally be the case in bioacoustics research where the spectrogram visualization metaphor remains the status quo. Instead, users assimilate collaboratively filtered interpretations from previous interactions, mediated by a feature learning model, allowing the information content depicted on the device to identify novel clusters of information associated with the original input.

4 Continuing Research

Our iterative design of the Mephistophone is ongoing, with subsequent versions informed by the efficacy of our sensing, actuating, illuminating and mapping models. As bioacoustic source material we have been using a subset of an open source database of bird calls available from xeno-canto¹ released as part of the BirdCLEF2014 competition². Two autonomous embedded micro-controllers interact with various features from the audio: the first extracting the spectrum of the signal and passing the output to the actuators informing the resultant spectrogram on the physical model of the surface; the second receiving the output of a timbral analysis of the audio and passing it to both projection and

¹ <http://www.xeno-canto.org>.

² <http://www.imageclef.org/2014/lifeclef/bird>.

embedded illumination sources. This mode is sufficiently fast and robust to run the analysis either on prerecorded audio, as currently described, or in real-time on the local soundscape; the latter for the time being suffers from significant internal noise within the machine, generating haptic/acoustic feedback loops. In informing learned gestures mapped to acoustic features, the Mephistophone takes previous users' sensed gestures as input, building a collaboratively filtered set of gestures correlated to a training set of acoustic samples. Users' gestural interactions with the system are collected in response to a prerecorded set of sounds and subsequent actuation on the surface as a result of a given audio seed are a composite of a mapping of the raw audio features extracted from the sample, and the collaboratively filtered behavior of prior users given the same seed.

'An enabling technology to build shape changing interfaces through pneumatically-actuated soft composite materials' [25] offers a soft robotics approach to shape change without the limitations of our arrangement of actuators thus allowing more complex shape and gesture dynamics. These types of changes are enumerated as 'orientation, form, volume, texture, viscosity, spatiality, adding/subtracting, and permeability' [18,25]. Within this space, the Mephistophone currently offers interaction in the first three instances. Additionally, we are exploring the potential of comparing the mapping of conventional spectral and timbral feature versus a learned set of acoustic features to the physical and illuminated display surface. Using feature learning approaches on the audio alone, the mapping of the raw signal to the TUI can reflect a higher dimensional representation than afforded by traditional displays and perhaps a more intuitive mapping than afforded by conventional acoustic features optimized for 2D displays [13]. In our next implementation, we are exploring the use deep belief models for generating learned features from a composite of audio and collaboratively filtered prior gestures. This will be used to construct a system wherein subsequent information from environmental soundscapes is augmented by inference from past interactions.

5 Discussion

Through ongoing research combining collaborative filtering, feature learning and human inference tasks we are exploring the haptic and visual affordances of a 2.5D TUI for bioacoustics using collaborative inference models. We are exploring novel representations in the audio representation-space and how a transition from spectral to timbral visualization can enhance user cognition. Despite the varied information, be it timbral, spectral or learned, available when treating the Mephistophone as a visualizer, preliminary user testing has proposed that users maintain sufficient familiarity with the information mapped to the surface of the Mephistophone to consistently, if arbitrarily, define regions of interest for given pre-recorded sounds or *pathways of interest* across soundscapes [24]. In further mixed-initiative development, users will tangibly interact with regions or pathways of interest collaboratively filtered by past users mediated by the

Mephistophone, yielding filtered output represented by the selection. As an installation piece, the Mephistophone collects users' literal impressions delineating perceived regions and pathways of interest informed on the surface. Through analysis of deep belief outputs from collaboratively filtered behaviors a learned mapping from a data set of prerecorded environmental soundscapes can be used as the basis for a generative model of subsequent user interactions. This in turn provides feedback to the algorithms refining the boundaries of regions of interest in a given signal resulting from learned latent interaction classes. As subsequent users define regions or pathways of interest, the embodied cognition of the Mephistophone can haptically reinforce learned timbre-space pathways. By training algorithms to reflect prior users' decision regarding regions and pathways of interest we provide a novel mechanism for subsequent users to be trained by, rather than on, the Mephistophone, augmenting their understanding of soundscapes old and new and further simplifying the task of training novice users to assess new recordings.

References

1. Caclin, A., McAdams, S., Smith, B.K., Winsberg, S.: Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. *J. Acoust. Soc. Am.* **118**(1), 471–482 (2005)
2. Dand, D., Hemsley, R.: Obake: interactions on a 2.5d elastic display. In: *Proceedings of the Adjunct Publication of the 26th Annual ACM Symposium on User Interface Software and Technology*, pp. 109–110. *UIST 2013 Adjunct*. ACM, New York, NY, USA (2013). <http://doi.acm.org/10.1145/2508468.2514734>
3. Dennis, J., Tran, H.D., Li, H.: Spectrogram image feature for sound event classification in mismatched conditions. *IEEE Sig. Process. Lett.* **18**(2), 130–133 (2011)
4. Depaetere, M., Pavoine, S., Jiguet, F., Gasc, A., Duvail, S., Sueur, J.: Monitoring animal diversity using acoustic indices: implementation in a temperate woodland. *Ecol. Ind.* **13**(1), 46–54 (2012)
5. Edge, D., Blackwell, A.: Correlates of the cognitive dimensions for tangible user interface. *J. Vis. Lang. & Comput.* **17**(4), 366–394 (2006)
6. Farina, A., Pieretti, N., Piccioli, L.: The soundscape methodology for long-term bird monitoring: a mediterranean europe case-study. *Ecol. Inform.* **6**(6), 354–363 (2011)
7. Hook, J., Taylor, S., Butler, A., Villar, N., Izadi, S.: A reconfigurable ferromagnetic input device. In: *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology*, pp. 51–54. ACM (2009)
8. Jorda, S., Kaltenbrunner, M., Geiger, G., Bencina, R.: The reactable*. In: *Proceedings of the International Computer Music Conference (ICMC 2005)*, Barcelona, Spain, pp. 579–582 (2005)
9. Kuhn, G.: Description of a color spectrogram. *J. Acoust. Soc. Am.* **76**(3), 682–685 (1984)
10. Leithinger, D., Ishii, H.: Relief: a scalable actuated shape display. In: *Proceedings of the Fourth International Conference on Tangible, Embedded, and Embodied Interaction*, pp. 221–222. ACM (2010)
11. McAdams, S.: Perspectives on the contribution of timbre to musical structure. *Comput. Music J.* **23**(3), 85–102 (1999)

12. Menon, V., Levitin, D., Smith, B.K., Lembke, A., Krasnow, B., Glazer, D., Glover, G., McAdams, S.: Neural correlates of timbre change in harmonic sounds. *Neuroimage* **17**(4), 1742–1754 (2002)
13. Mohamed, A.R., Dahl, G.E., Hinton, G.: Acoustic modeling using deep belief networks. *IEEE Trans. Audio, Speech, Lang. Process.* **20**(1), 14–22 (2012)
14. Obrist, M.K., Pavan, G., Sueur, J., Riede, K., Llusia, D., Márquez, R.: Bioacoustics approaches in biodiversity inventories. *Abc Taxa* **8**, 68–99 (2010)
15. Pieretti, N., Farina, A., Morri, D.: A new methodology to infer the singing activity of an avian community: the acoustic complexity index (aci). *Ecol. Indic.* **11**(3), 868–873 (2011)
16. Piper, B., Ratti, C., Ishii, H.: Illuminating clay: a 3-d tangible interface for landscape analysis. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 355–362. ACM (2002)
17. Purchase, H.C., Andrienko, N., Jankun-Kelly, T.J., Ward, M.: Theoretical foundations of information visualization. In: Kerren, A., Stasko, J.T., Fekete, J.-D., North, C. (eds.) *Information Visualization*. LNCS, vol. 4950, pp. 46–64. Springer, Heidelberg (2008)
18. Rasmussen, M.K., Pedersen, E.W., Petersen, M.G., Hornbæk, K.: Shape-changing interfaces: a review of the design space and open research questions. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 735–744. ACM (2012)
19. Segnini, R., Sapp, C.: Scoregram: displaying gross timbre information from a score. In: Kronland-Martinet, R., Voinier, T., Ystad, S. (eds.) *CMMR 2005*. LNCS, vol. 3902, pp. 54–59. Springer, Heidelberg (2006)
20. Sequera, R.S.: Timbrescape: a musical timbre and structure visualization method using tristimulus data. In: *Proceedings of the 9th International Conference on Music Perception and Cognition (ICMPC)*, Bologna (2006)
21. Siedenbarg, K.: An exploration of real-time visualizations of musical timbre. In: *Welcome to the 3rd International Workshop on Learning Semantics of Audio Signals*, p. 17 (2009)
22. Sueur, J., Pavoine, S., Hamerlynck, O., Duvail, S.: Rapid acoustic survey for biodiversity appraisal. *PLoS One* **3**(12), e4065 (2008)
23. Tillmann, B., McAdams, S.: Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis) similarities. *J. Exp. Psychol. Learn. Mem. Cogn.* **30**(5), 1131 (2004)
24. Wollner, P.K.A., Herman, I., Pribadi, H., Impett, L., Blackwell, A.F.: Mephistophone. Technical Report UCAM-CL-TR-855, University of Cambridge, Computer Laboratory, June 2014. <http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-855.pdf>
25. Yao, L., Niiyama, R., Ou, J., Follmer, S., Della Silva, C., Ishii, H.: Pneui: pneumatically actuated soft composite materials for shape changing interfaces. In: *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, pp. 13–22. ACM (2013)