Algorithmic Spatialization Using Object-Based Audio and Indoor Positioning System

Yuan-Yi Fan

Leonardo Music Journal, Volume 29, 2019, pp. 25-30 (Article)

Published by The MIT Press

Algorithmic Spatialization Using Object-Based Audio and Indoor Positioning System

YUAN-YI FAN

ABSTRACT

The author presents a novel compositional framework to guide designing interplay between moving listeners and sound objects in space. Demonstrated by a case study of interactive octophonic installation, the presented framework offers new ways to articulate and analyze artistic interplay using real-world location context as a spatial composition canvas.

ORGANIZED SOUND IN SPACE

Using sound as an artistic medium is a creative practice in contemporary life [1], and spatial practices in sound art involve techniques to organize sound in space [2] and understand sound in relation to place [3]. Historically, key spatialization research in electronic music can be traced back to the 1950s [4]. At EXPO '58 in Brussels, Edgard Varèse's *Poème électronique* and Iannis Xenakis's *Concret PH* for the Philips Pavilion were examples of articulating space through sound. More spatial sound projection systems were seen at EXPO '70 in Osaka, including Karlheinz Stockhausen's *Expo* at the German Pavilion, Iannis Xenakis's *Hibiki-Hana-Ma* at the Japanese Pavilion and David Tudor's *Microphone* at the Pepsi Pavilion [5,6]. Sound, space and listeners are common elements of interplay in this research.

STRATEGIES FOR ARTICULATING SPACE USING SOUND

In spatiomusical composition strategies [7] and computer-aided composition [8], spatial elements are treated as compositional materials for developing music structures in space, oftentimes under the assumption of stationary listeners. Although topics of nonstationary listeners are rarely the focus in researches that address interactions between composition techniques and spatial audio technologies [9,10], they need to be addressed in interactive cases where audience participation and interactivity play a central role. To investigate the topics of nonstationary listeners in a spatiomusical composition context, research in formal music representations [11] provides inspirations in developing generative strategies for machine improvisation [12] and algorithmic spatialization [13]. On the other hand, studies of spatial concepts [14], spatial attributes [15] and spatial schemata [16] provide a set of analytical perspectives to understand and characterize sound in relation to space.

Yuan-Yi Fan (artist, researcher). Email: yyf@yuanyifan.com.
Web: www.yuanyifan.com.

See www.mitpressjournals.org/toc/lmj/29 for supplemental files associated with this issue.

TECHNOLOGIES FOR OBJECT-BASED AUDIO AND INDOOR POSITIONING SYSTEM

The ability to capture spatial relations of listeners to space is critical in developing the idea of using real-world location context as a spatial composition canvas. In spatial audio production, object-based audio differs from traditional channel-based audio in its potential for interactive scenes. It provides parametric controls of spatial-temporal models of sound sources and independence from the reproduction setup [17], which are practical for sound art use cases. Higher-order ambisonics (HOA) and Vector Base Amplitude Panning (VBAP) are two common spatial audio reproduction techniques that are briefly discussed in the case study section. While there is some industry-driven research [18] and the upcoming MPEG standard [19] that facilitates object-based audio for more listener involvement and for richer media experiences, existing product use cases are limited to static listening situations in cinemas and living rooms, as opposed to more dynamic situations in sound art.

As commercial IPS (indoor positioning system) solutions become more affordable, microlocation and spatial context of moving listeners could become potential abstract compositional materials. IPS uses various wireless technologies, including GPS-based, RFID, cellular-based, UWB, WLAN, Bluetooth and others [20]. A Bluetooth low-energy beacon, such as Estimote [21], is one of the affordable IPS solutions capable of providing real-time indoor location, orientation and motion data. While recent research addresses crowd-based music interactivity and composition strategies using mobiles, sensors and robotics [22–25], relatively few investigate the case of nonstationary listeners systematically.

 doi:10.1162/LMJ_a_01058

Related to composing interactive soundscapes using GPS data [26,27], this article discusses new ways of composing interactive soundscape using IPS data.

THE CHALLENGES

Designing interplay between moving listeners and sound objects in space poses aesthetic and technical challenges to electroacoustic composers and sound artists. While there is important research addressing interactions and tensions between spatial composition techniques and spatial audio technologies, a framework that facilitates articulating and analyzing artistic interplay between moving listeners and sound objects in space is missing. To address these challenges, I propose a novel compositional framework consisting of object-based audio, indoor positioning system and algorithmic spatialization strategies. The proposed framework is both conceptual and technological. Conceptually, it aims to provide critical dimensions that enable analytical discussion of spatial development and interactivity with regard to moving listeners. Technologically, it aims to investigate necessary tools that facilitate composing spatial forms and interactions responsive to moving listeners. I demonstrate the presented framework by a case study of interactive octophonic installation.

CASE STUDY—SYMBOLS AND BOUNDARIES

An artistic statement of this case study is inherent in its composition plan. Instead of using a result-oriented compositional approach, the artistic interplay is steered by the composition plan that organizes indoor location data, operates on spatial representations and renders spatial sounds. Indeterminacy introduced by delegating trajectory creation of sound objects to moving listeners makes it an interactive ecosystem, the sonic output of which favors aleatory and emergent spatial development. Sound becomes a spatial interface between autonomous DSP modules and the external environment, where the spatial structural development is a result of previous interaction patterns. In this ecosystem, the computer observes moving listeners in the environment, alters its internal representation for algorithmic spatialization and reacts to spatial movement history by projecting a heatmap-like soundscape. In this sense, this ecosystem exhibits a kind of spatial memory.

Central to this case study, the composition plan consists of three sonic layers (Fig. 1). The always-on layer invites nearby spectators to experience the installation by looping ambient sound objects in space. In terms of interactivity, the static spatial structure in this sonic layer provides an initial condition of a human-machine system. Spectators become potential listeners once they walk into the installation space.

The responsive layer is responsible for the localized auditory icons. When a spectator enters the installation space, the location beacons detect an entrance event that triggers an auditory icon playback at the same location as the entrance event. This marks the birth of a sound object. When the listener exits the installation space, a different auditory icon is played at the exit location. This marks the death of the sound object. Localized sonic events in this layer make listeners aware of the boundary of this human-machine system.

The evolutionary layer loops sound objects in space based on algorithmically generated trajectories. Indoor location data of moving listeners are used to create trajectories for sound objects and to steer spatial structure development over time. Figure 2 illustrates an example computational process that steers spatial structural development of sound objects based on accumulated location data of moving listeners. Similar to the heatmap technique in data visualization, a virtual canvas recomputes the representation of location data at each entrance-exit interval. Spatial and statistical analysis of the virtual representation steers sound spatialization by playing low pulse drone sounds at less popular zones, i.e. cells with value = 1, and by looping percussive sounds along hot spots, i.e. cells with value = 2. Spatial development in this layer encourages listeners to explore the installation space and thus generates more spatial data to iteratively enrich the spatial development.

Zirkonium spatialization software [28] supports OSC

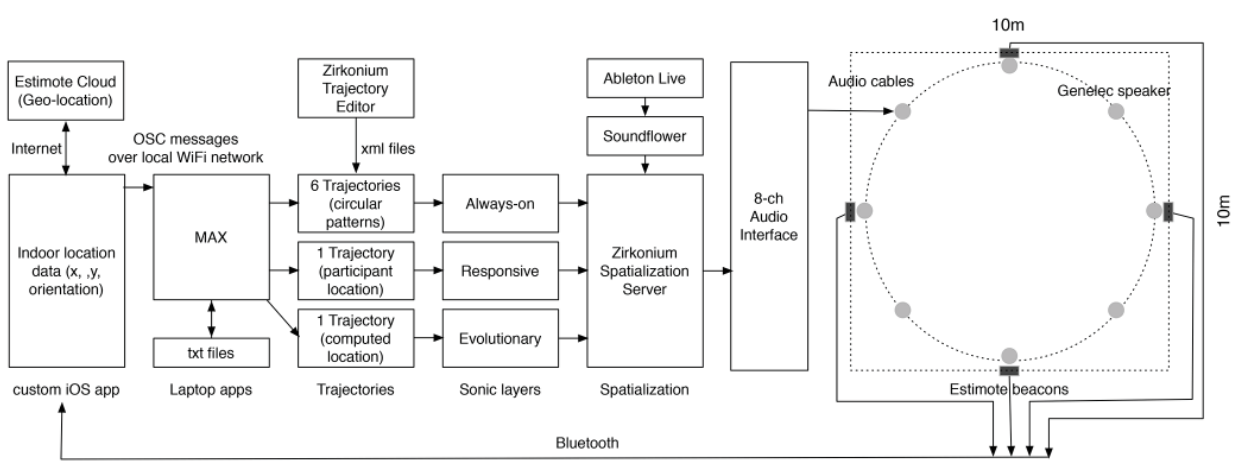

Fig. 1. A technological overview of the case study. (© Yuan-Yi Fan)

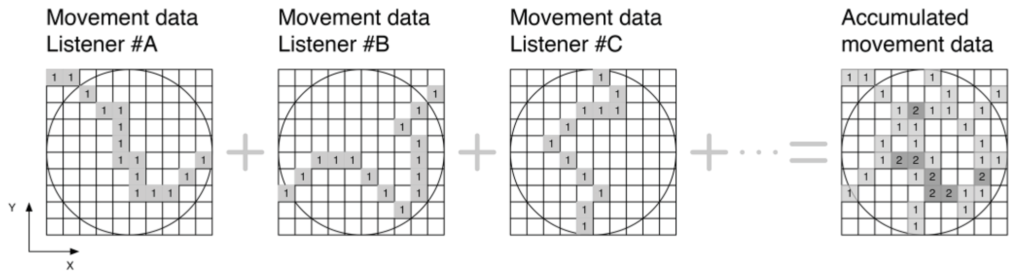

Fig. 2. A computational process that maps indoor location data to a heatmap-like spatialization scheme. (© Yuan-Yi Fan)

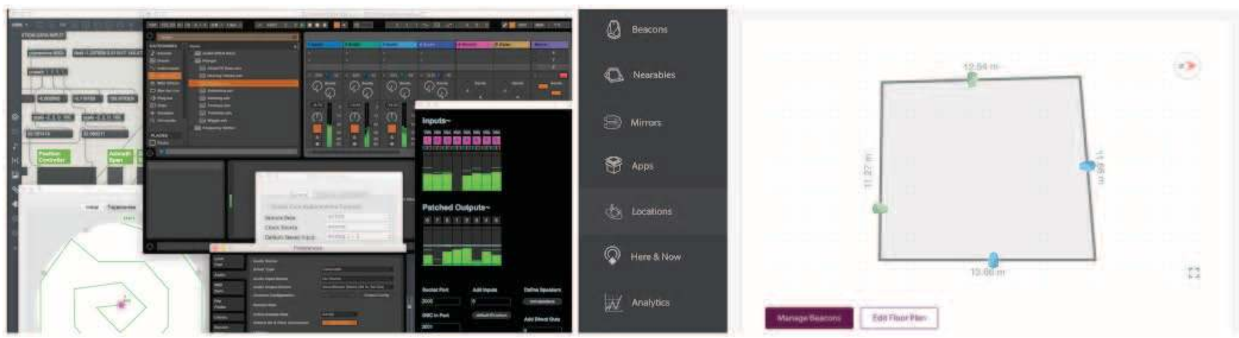

Fig. 3. A snapshot of the workflow using Zirkonium, Max, Ableton and Estimote API. (© Yuan-Yi Fan)

Fig. 4. The indoor positioning system consists of one iOS device and four Estimote sensors. (© Yuan-Yi Fan)

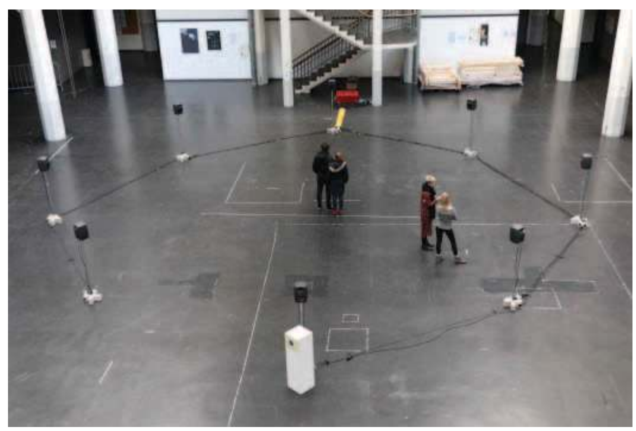

Fig. 5. A top-down view of the 10-square-meter octophonic installation space. (© Yuan-Yi Fan)

(OpenSoundControl) [29] and is therefore used for interactive controls and spatial composition in the case study (Fig. 3). While both VBAP and HOA are supported in Zirkonium, I use HOA for playback smoothness at the cost of localization accuracy. Estimote beacons (Fig. 4) and its cloud API are used as an IPS solution that enables real-time location tracking of moving listeners. Coordinates and orientation data of the listeners (x, y, theta) are captured using Estimote iOS Indoor Location SDK, transmitted over local area network in OSC format and stored on a laptop for trajectory computation.

Physical dimensions of this sound installation are determined primarily based on Estimote spatial sensing resolution with considerations of HOA localization accuracy. At the time of this installation, the best spatial sensing resolution using four Estimote sensors was 1 × 1 m in a 10 × 10 m area (Fig. 5). To the best of my knowledge, there are perceptual studies of first-order and higher-order ambisonics at central and off-center listening positions, but no systematic perceptual study at dynamic listening positions. For a 5th-order HOA, the best-achieved localization accuracy for a central listening position is 3 degrees using 12 loudspeakers with a circumradius of 3.86 m [30]. A lower-localization accuracy is expected in the case study, since only the 3rd-order HOA is used in the octophonic setup, with a circumradius of 5.94 m.

Technical realization of the proposed framework allows us to design interplay using various types of spatial interactions and conceptual constructs. The types of spatial interactions enabled by the proposed framework include localized auditory icons for event notification, dynamic creation of sound object trajectories using listeners' movement data and algorithmic spatialization as auditory display. These types of interaction also help demonstrate gaps in creating auditory experience for mixed reality [31] using research compositional tools [32] versus using commercial production tools such as Unity, iOS ARKit and Facebook Spatial Workstation.

This framework offers three conceptual constructs to guide designing interplay between moving listeners and sound objects in space. First, localized sound objects can be architected to provide hierarchy, grouping, layering [33] and spatial affordance [34] in an auditory interface. Second, spatial interactions can be modeled as structural elements in an ecosystemic approach [35] or as low-level processes of a multiscale approach [36] to algorithmic spatialization. Third, spatial interaction data can be decomposed and organized as a formal input representation to a variety of musical automata [37] for algorithmic spatialization. These constructs encourage future creative machine intelligence research at the intersection of MIDI [38], SpatDif [39], graphical score [40], action and movement notation [41] and symbolic graphical notation [42]. For example, symbolic operations on an acoustic scene representation could provide a basis for machine spatiomusical improvisation and location-based sonic narratives.

ANALYSIS OF THE CASE STUDY

I exhibited the case study at the 2015 inSonic festival, the theme of which focused on aesthetic concepts of spatial audio between music and sound art [43]. The case study demonstrates a compositional use of moving listeners as spatial elements [44] in the context of sound installation art. It extends one of the common spatial composition techniques [45] from predefined trajectories to dynamic trajectory creation using moving listeners' indoor location data. Instead of using symbolic elements for computer-aided composition and spatialization [46], the use of spatial representations provides a virtual canvas for rule-based computational operations. Similar to the graphical score approaches [47], such virtual canvas allows computation of spatial structures and development using real-world location data. The presented framework investigates necessary tools that could facilitate experiments in space-form investigation and acousmatic image viewed from a moving listener's vantage point [48].

As a conceptual framework, synthesis of techniques and strategies from related works provides a working set of dimensions for analyzing interplay in sound art installation. These dimensions include spatial elements [49], spatial development [50], computer-aided composition [51], formal representation [52], symbolic processes [53], types of automaton [54], algorithmic composition [55], generative strategies [56], algorithmic spatialization [57], spatial structure [58], spatial concepts [59], spatial attributes [60] and spatial schemata [61].

Technical limitations of the presented system are addressed in terms of temporal and spatial sampling resolutions. Temporal resolution of the Estimote sensors is limited by their minimal advertising interval, i.e. 100 ms. Spatial resolution provided by the four Estimotes is limited to a 1 square meter grid, which gives at least 88 spatial sampling points for our circular installation space. While an objective assessment is needed for evaluating the third-order HOA localization accuracy in the octophonic setup, a comparative estimate is addressed in the case study section.

Shifting from a composer's perspective to a listener's perspective, perceptions of spatial elements and structural developments depend on attributes of space, effectiveness of spatialization techniques and spatial relations between listeners and loudspeakers. Based on the visitors' feedback, both the always-on and the responsive sonic layers provide effective spatial auditory cues in terms of my conceptual intents. However, spatialization results from the evolutionary layer become less effective when the accumulated number of visitors exceeds a certain threshold or when more than one visitor is interacting with the system at the same time. The threshold varies based on cumulative patterns of the listeners' movement. Although one can set the threshold value to an arbitrarily small number to avoid ineffective spatialization cues, that limits the length of intended algorithmic spatial development.

SUMMARY

This article presents a novel compositional framework that consists of object-based audio, indoor positioning system and algorithmic spatialization strategies. Conceptually, it offers means to articulate artistic interplay between moving listeners and sound objects using real-world location context as a spatial composition canvas. It also provides a working set of dimensions for analyzing interplay in sound art installation. Technologically, it investigates necessary tools to facilitate these conceptual explorations. The presented use case demonstrates technical realization of these explorations, which include three types of spatial interactions and three conceptual constructs. Aesthetically, the case study considers moving listeners and sound objects as one interactive ecosystem where its sonic output favors aleatory and emergent spatial development and exhibits a kind of spatial memory by projecting a heatmap-like soundscape. Beyond sound art, the case study suggests that this framework has great creative potentials for machine spatiomusical improvisation and mixed-reality audio applications. These potentials should be further investigated to outline needs for new transmission specifications, model representations, evaluation methods and design tools that would better integrate sound field synthesis and listener location–tracking workflows.

Acknowledgments

The author wishes to thank Curtis Roads for research guidance and the inSonic team for equipment support and exhibition production.

References and Notes

1 Nicolas Collins, "Introduction," *Leonardo Music Journal* **23** (2013) pp. 1–2.

2 Yuan-Yi Fan, and David Minnen, "*Move That Sound There*: Exploring Sound in Space with a Markerless Gestural Interface," *Leonardo Music Journal* **23** (2013) pp. 31–32.

3 Gascia Ouzounian, "Sound Art and Spatial Practices: Situating Sound Installation Art Since 1958," Diss. UC San Diego, 2008.

4 Curtis Roads and John Strawn, *The Computer Music Tutorial* (Cambridge, MA: MIT Press, 1996).

5 Curtis Roads, *Composing Electronic Music: A New Aesthetic* (Oxford Univ. Press, 2015).

6 Billy Klüver, Julie Martin and Barbara Rose, eds., *Pavilion: Experiments in Art and Technology* (Boston: EP Dutton, 1972).

7 Natasha Barrett, "Spatio-Musical Composition Strategies," *Organised Sound* **7**, No. 3, 313–323 (2002).

8 Marlon Schumacher and Jean Bresson, "Spatial Sound Synthesis in Computer-Aided Composition," *Organised Sound* **15**, No. 3, 271–289 (2010).

9 Marije AJ Baalman, "Spatial Composition Techniques and Sound Spatialisation Technologies," *Organised Sound* **15**, No. 3, 209–218 (2010).

10 Nils Peters, Georgios Marentakis and Stephen McAdams, "Current Technologies and Compositional Practices for Spatialization: A Qualitative and Quantitative Analysis," *Computer Music Journal* **35**, No. 1, 10–27 (2011).

11 Roads and Strawn [4]; Robert Rowe, *Machine Musicianship* (Cambridge, MA: MIT Press, 2004).

12 Gérard Assayag, "Creative Symbolic Interaction," *40th International Computational Musical Conference and 11th Sound and Music Computer Conference* (ICMC/SMC joint conf.), ICMA, SMC, National and Kapodistrian University of Athens, IRMA, 2014.

13 Alex McLean and Roger T. Dean, eds., *The Oxford Handbook of Algorithmic Music* (Oxford, U.K.: Oxford Univ. Press, 2018).

14 Denis Smalley, "Space-Form and the Acousmatic Image," *Organised Sound* **12**, No. 1, 35–58 (2007).

15 Francis Rumsey, "Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm," *Journal of the Audio Engineering Society* **50**, No. 9, 651–666 (2002).

16 Gary S. Kendall and Mauricio Ardila, "The Artistic Play of Spatial Organization: Spatial Attributes, Scene Analysis and Auditory Spatial Schemata," *International Symposium on Computer Music Modeling and Retrieval* (Berlin, Heidelberg: Springer, 2007).

17 Frank Melchior and Sascha Spors, "Spatial Audio Reproduction: from Theory to Production," tutorial, 129th Convention of the AES, San Francisco, CA, U.S.A., 2010.

18 Robert Bleidt et al., "Object-Based Audio: Opportunities for Improved Listening Experience and Increased Listener Involvement," *SMPTE Motion Imaging Journal* **124**, No. 5, 1–13 (2015).

19 Jonas Engdegård et al., "Spatial Audio Object Coding (SAOC)—The Upcoming MPEG Standard on Parametric Object Based Audio Coding," *Proceedings: 124th AES Convention* (Amsterdam: Audio Engineering Society, 2008).

20 Hui Liu et al., "Survey of Wireless Indoor Positioning Techniques and Systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **37**, No. 6, 1067–1080 (2007).

21 www.estimote.com (accessed 10 February 2019).

22 Golon Levin, Scott Gibbons and Gregory Shokar, *Dialtones: A Telesymphony*. Staalplaat, 2005. Compact disc.

23 Mark Feldmeier and Joseph A. Paradiso, "An Interactive Music Environment for Large Groups with Giveaway Wireless Motion Sensors," *Computer Music Journal* **31**, No. 1, 50–67 (2007)

24 Jason Freeman, "Large Audience Participation, Technology, and Orchestral Performance," *International Computer Music Conference Proceedings, 2005* (Barcelona: International Computer Music Association, 2005)

25 Edwin van der Heide, "Spatial Sounds (100dB at 100km/h) in the Context of Human Robot Personal Relationships," *Third International Conference on Human-Robot Personal Relationships* (Berlin, Heidelberg: Springer, 2010).

26 Atau Tanaka and Petra Gemeinboeck, "A Framework for Spatial Interaction in Locative Media," *Proceedings of the 2006 Conference on New Interfaces for Musical Expression* (Paris: IRCAM—Centre Pompidou, 2006) pp. 26–30.

27 Rob van Rijswijk and Jeroen Strijbos, "Sounds in Your Pocket: Composing Live Soundscapes with an App," *Leonardo Music Journal* **23** (2013) pp. 27–29.

28 David Wagner et al., "Introducing the Zirkonium MK2 System for Spatial Composition," *Proceedings ICMC/SMC/2014* (Athens: International Computer Music Association, 2014) pp. 823–829.

29 Matthew Wright, "Open Sound Control: An Enabling Technology for Musical Networking," *Organised Sound* **10**, No. 3, 193–200 (2005).

30 Hagen Wierstorf, Alexander Raake and Sascha Spors, "Assessing Localization Accuracy in Sound Field Synthesis," *The Journal of the Acoustical Society of America* **141**, No. 2, 1111–1119 (2017).

31 Paul Milgram and Fumio Kishino, "A Taxonomy of Mixed Reality Visual Displays," *IEICE TRANSACTIONS on Information and Systems* **77**, No. 12, 1321–1329 (1994).

32 Wagner et al. [28]; Miller S. Puckette, "Pure Data" *Proceedings of the 1997 International Computer Music Conference* (Thessaloniki, Greece: International Computer Music Association, 1997); Tobias Höllerer, JoAnn Kuchera-Morin and Xavier Amatriain, "The Allosphere: A Large-Scale Immersive Surround-View Instrument," *Proceedings of the 2007 Workshop on Emerging Displays Technologies: Images and Beyond: The Future of Displays and Interaction* (San Diego, CA: ACM, 2007).

33 Albert S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (Cambridge, MA: MIT Press, 1994).

34 William W. Gaver, "Technology Affordances," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montreal: ACM, 1991).

35 Agostino Di Scipio, "'Sound Is the Interface': From Interactive to Ecosystemic Signal Processing," *Organised Sound* **8**, No. 3, 269–277 (2003).

36 Roads [5].

37 Joel Chadabe, *Electric Sound: The Past and Promise of Electronic Music* (Pearson, 1997).

38 Li-Chia Yang, Szu-Yu Chou and Yi-Hsuan Yang, “MidiNet: A Convo-Lutional Generative Adversarial Network for Symbolic-Domain Music Generation,” arXiv preprint, arXiv:1703.10847 (2017).

39 Nils Peters, Sean Ferguson and Stephen McAdams, “Towards a Spatial Sound Description Interchange Format (SpatDIF),” *Canadian Acoustics* **35**, No. 3, 64–65 (2007).

40 Iannis Xenakis, *Formalized Music: Thought and Mathematics in Composition* (Hillsdale, NY: Pendragon Press, 1992).

41 Dieter Daniels and Sandra Naumann, eds., *See This Sound: Audiovisuology Essays 2: Histories and Theories of Audiovisual Media and Art* (Cologne: Walther König, 2011).

42 Rama Gottfried, “SVG to OSC Transcoding: Towards a Platform for Notational Praxis and Electronic Performance,” *Proceedings of the First International Conference on Technologies for Notation and Representation* (Paris: TENOR, 2015).

43 www.zkm.de/en/event/2015/11/globale-insonic2015 (accessed 10 February 2019).

44 Barrett [7].

45 Baalman [9].

46 Schumacher and Bresson [8].

47 Roads [5]; Xenakis [40].

48 Smalley [14].

49 Barrett [7].

50 Baalman [9].

51 Schumacher and Bresson [8].

52 Roads and Strawn [4]; Rowe [11].

53 Rowe [11].

54 Roads and Strawn [4]; Rowe [11].

55 Roads and Strawn [4]; Roads [5]; Assayag [12]; McLean and Dean [13].

56 Roads [5].

57 McLean and Dean [13].

58 Smalley [14].

59 Smalley [14].

60 Rumsey [15].

61 Kendall and Ardila [16].

Manuscript received 11 November 2018.

YUAN-YI FAN *studied computer music, media arts and multimedia engineering in the Media Arts and Technology program at University of California, Santa Barbara, where he received his PhD in 2015. His research investigates collective expression by creating systems that augment audience participation and multimodal interactivity. He currently leads audio research and engineering at a startup in Los Angeles. Before attending UCSB, he built MEMS ultrasound transducer arrays at the Ultrasound Imaging Laboratory in Taiwan.*