Survey and Implications for the Design of New 3D Audio Authoring Tools

Justin Mathew, Stéphane Huot, Brian F. G. Katz

▶ To cite this version:

HAL Id: hal-01517188

https://hal.inria.fr/hal-01517188

Submitted on 27 Aug 2018

Survey and Implications for the Design of New 3D Audio Authoring Tools

Justin D. Mathew · Stéphane Huot · Brian FG Katz

Abstract 3D audio production tools vary from low-level programming libraries to higher-level user interfaces that are used across a wide range of applications today. However, many of these user interfaces are underdeveloped with limited functionality, forcing users to resort to ad hoc solutions with other tools or programming languages. Identifying these limitations and custom methods are needed to inform the development of new user interfaces. Towards this end, an on-line survey was conducted with current practitioners to gather ethnographic information on their tools, methods, and opinions. Results of the survey identified specific methods and limitations regarding *Audio Rendering*, *Visual Feedback*, *Functionality*, and *Workflow Integration*. These results also shed light on three basic tasks that have to be performed interactively with 3D audio production tools: *Defining the Rendering Space*, *Creation and Manipulation of Audio Objects*, and *Use of Feedback*. This classification helps organize the creative needs for 3D audio tools that address issues within the workflow and low-level functionality of systems.

Justin D. Mathew
Inria Lille-Nord Europe, LIMSI-CNRS,
Université Paris-Saclay
Parc scientifique de la Haute-Borne
40 avenue Halley - Bât A, 59650 Villeneuve-d'Ascq, France
Tel.: +123-45-678910, E-mail: justin.mathew@inria.fr

Stéphane Huot
Inria Lille-Nord Europe
Parc scientifique de la Haute-Borne
40 avenue Halley - Bât A, 59650 Villeneuve d'Ascq, France
Tel.: +33 (0)3 59 35 87 89, E-mail: Stephane.Huot@inria.fr

Brian FG Katz
Sorbonne Universités, UPMC Univ Paris 06, CNRS,
Institut d'Alembert
4, place Jussieu, Paris, France
E-mail: brian.katz@upmc.fr

1 Introduction

Immersive environments have significantly grown in popularity and are becoming more affordable for consumers. In the case of digital audio, producing a sense of an aural space can give listeners realistic and imaginative experiences for a variety of purposes across different domains: music (recorded or performed), installations for sonic art, cinema, TV and video games. To produce such content, users create audio data through recording or sonification techniques (e.g., musical parameters, sound synthesizers, etc.), which they manipulate through signal processing techniques (e.g., volume, EQ, etc.) to set their basic audio parameters, such as loudness or frequency content. A sense of 3D spaciousness can be added to the production through rendering audio data with advanced 3D rendering methods and reproduction setups, which allow users to control the spatialization effect. Traditionally, this is done with standard mixing interfaces that pan the audio between two speakers, but recent setups have been augmented with renderers and input devices with more Degrees of Freedom (DoF) that should give better control over the 3D space of the audio production. However, many of these systems are considered impractical or cumbersome to use due to: (1) the complexity of setting up and configuring an appropriate tool-chain for 3D audio mixing that can suit the large variety of audio recording techniques, 3D audio formats, and rendering systems; (2) the lack of appropriate interactive tools to efficiently manipulate the high-dimensional data and the

3D acoustic characteristics of the audio (Peters et al, 2011; Mathew et al, 2014).

In more detail, audio spatialization is achieved through advanced recording, synthesis and rendering of acoustic spatial properties. Recording methods, such as the *Soundfield microphone*[1], encode spatial properties into the audio data which can be decoded onto various speaker configurations. Other methods, such as *Vector-Based Amplitude Panning (VBAP)* (Pulkki, 1997), compute a ratio of gains and/or filter the audio signal to specific output channels based on user-defined positional information of the audio and a predefined speaker configuration. In addition, *Binaural* rendering techniques process audio through the use of perceptual filters called Head-Related Transfer Functions (HRTF) for reproduction on headphones (Kendall, 1995). These methods create the impression that a sound is originating from a specific direction, resulting in a more spatial listening experience. From the user's perspective, these rendering techniques are intended to allow users to control high-level spatial parameters without worrying about the actual low-level signal processing of audio channels. Although this is suitable for basic speaker configurations (e.g. stereo, headphones, Surround Sound), it can become complicated with complex 3D speaker configurations since it requires a firm understanding of the acoustics and signal processing method of the renderers and precise routing of channels (Peters et al, 2011; Geier et al, 2010). In addition, user interfaces of the rendering tools are under-designed making them complex, tedious, and limited in functionality. Even though there is a significant increase of interest in this problem, which we will discuss in the next section, users still mention difficulties for interacting with the spatial parameters of 3D audio tools (Mathew et al, 2014; Gelineck and Korsgaard, 2015). Moreover, the variety of domains and productions methods makes it difficult to identify issues that range across all users. To address these issues and gather a better understanding of how practitioners produce spatial audio content, we conducted an on-line survey targeting professional and amateur users of 3D audio. We collected answers from participants about their projects, their production systems (hardware and software), how they author, notate, and evaluate spatial parameters, and their assessment of the advantages and limitations of those tools. Their responses helped us to identify methods and limitations in four major categories: *Audio Rendering*, *Visual Feedback*, *Functionalities*, and *Workflow Integration*. From these results, we identified high-level tasks that are required for 3D audio user interfaces, which we classify as *Defining the Rendering Space*, *Creation and Manipulation of Audio Objects*, and *Evaluation of the Listening Experience*. This classification provides a framework to inform the design of 3D audio user interfaces for more flexibility and functionality while addressing the needs of workflow for 3D audio production.

2 Studies on 3D Audio Authoring Tools

3D audio tools for consumers have been mostly introduced through plug-ins for Digital Audio Workstations (DAWs) and standard mixing consoles, but the use of programming languages and custom hardware are also popular choices. DAWs like Avid ProTools[2], Apple Logic Pro X[3], or Reaper[4] allow users to record multiple audio signals and mix them together through the bussing and summing of audio channels. Within this workspace environments, 3D audio plug-ins are used to control the spatial parameters and render the audio signals to the outputs according to the listening formats (e.g. mono, stereo, binaural, surround sound). Examples of available consumer plug-ins are the Spatial Audio Designer[5], Flux Ircam Spat[6], and the Dolby Atmos Panner[7] plug-ins (see Fig. 1(a)). They provide the ability to author, record, and edit the spatialization of audio interactively with the Graphical User Interfaces (GUI) of DAWs and plug-ins. Other tools provide external software to work alongside DAWs, such as the Dolby Atmos Monitoring Application[8] (see Fig. 1(b)), Sonic Emotion Wave 1 Designer (see Fig. 1(c)), and Performer[9] (see Fig. 1(d)). In addition, external hardware interfaces are available to aid users in controlling multiple parameters simultaneously. For example, new mixing consoles embed input devices such as trackballs[10] or dual joysticks[11] to augment the standard faders, knobs, and buttons. In addition, 3rd-party external input devices, such as the Behringer X-Touch Control Surface[12] and TouchOSC

---

[1] http://www.tslproducts.com/soundfield-type/soundfield-microphones/

[2] http://www.avid.com/pro-tools

[3] http://www.apple.com/logic-pro/

[4] http://www.reaper.fm/

[5] http://www.newaudiotechnology.com/en/products/spatial-audio-designer/

[6] http://www.fluxhome.com/products/plug_ins/ircam_spat-v3

[7] http://www.dolby.com/us/en/technologies/cinema/dolby-atmos.html

[8] http://www.dolby.com/us/en/technologies/cinema/dolby-atmos.html

[9] http://www2.sonicemotion.com/professional/

[10] http://solidstatelogic.com/studio/duality#&panel1-1

[11] http://www.ams-neve.com/products/post-production/dfc-gemini

[12] http://www.music-group.com/Categories/Behringer/Computer-Audio/Desktop-Controllers/X-TOUCH/p/P0B1X

iPad/iPhone application[13], allow users to map control parameters to spatial audio parameters in their software.

However, most of these tools are designed with respect to available listening formats within the DAW. For better flexibility, some low-level languages provide users with direct control over the rendering system, which enable a greater degree of freedom. Visual programming languages, such as Max[14], Pure Data[15], or OpenMusic[16] provide libraries of objects with built-in interactive tools that users can configure to their own needs. For example, the IRCAM Spat[17] or HOA Library[18] provide 3D audio rendering capabilities with the freedom to configure custom speaker setups and to customize methods for controlling spatial parameters. Textual programming libraries, such as OpenAL[19] or the Spatial Audio Toolbox for MATLAB[20], provide even lower-level functions to control renderers and spatial parameters. Even though these methods increase flexibility, they require a high level of expertise in acoustics, audio signal processing, and programming, which is actually not the case for most 3D audio content producers. Consequently, intermediate interactive systems are necessary that would both facilitate basic mixing thanks to a "low entry fee" (Wessel and Wright, 2002) and enable expert control of complex 3D audio setups.

Studying user interfaces and interaction techniques that are better adapted to audio spatialization is a relatively new topic, especially from an academic perspective. Efforts have mainly focused on comparing new ad hoc controllers with standard control methods for positioning a sound in a 3D virtual space. For instance, Carrascal and Jordà (2011) compared a standard mixing console with a multi-touch surface. In that study, participants completed the mixing tasks faster with the multi-touch interface and preferred it over the standard mixing interface. In another study, the majority of participants felt that the traditional mixing console was easier to use for the typical mixing task of positioning a sound, but that mid-air gestures gave them a better understanding of the 3D space (Gelineck and Korsgaard, 2015) even though there were mixed opinions about the benefits of mid-air interaction for creativity in a previous study (Churnside et al, 2011). A similar

comparison between a desktop mouse and a haptic device for creating trajectories of sound sources in a 3D space revealed better predictability and manageability with a mouse even though participants preferred the haptic device overall (Melchior et al, 2013). Such studies validate and evaluate specific design choices over traditional mixing methods, but they target a specific task and fall short in addressing other user needs for 3D audio production.

Ethnographic and exploratory studies can provide this understanding of users' needs and of their working process, which could ultimately inspire the design of better tools and techniques (Mackay, 2004). In addition, this approach can have a wider impact than the design of a unique technique, providing more general insights into the whole process. For instance, interviews with composers identified a need for temporal visual feedback of the spatial scene in 3D audio composition, which informed the design of new visualization and interaction techniques for composing spatial audio trajectories. Rather than comparing their prototypes with traditional techniques, the authors observed sessions with expert users and noted suggestions on how to improve their design ideas (Garcia et al, 2015; Favory et al, 2015). In an online survey of composers for 3D audio, Peters et al (2011) found that technical considerations, such as the need for real-time audio rendering and 3D visual representation, were highly desired functionalities for composers. Boutard and Guastavino (2012) followed a similar methodology with an online survey to gather the needs for archiving spatial audio production for future performances. This led to the development of a framework that classifies required information into technical skills and specifications, mapping strategies, characteristics of the performance venue, and perceptual effects to perform a spatial audio composition. Finally, interviews have also been used with professional recording and mixing engineers to understand their methodologies. This helped a morphological analysis of 3D audio tool resulting in a conceptual framework to help better define audio object properties and their associated control methods (Mathew et al, 2014). However, this study is limited to a few interviewees and focused closely on audio mixing rather than 3D audio production as a whole.

Building upon these works and methodologies, our online survey gathered information on current tools and techniques used by 3D audio users, with the overall objective to better understand and generalize the limitations of state-of-the-art interactive tools for 3D audio production.

---

[13] http://hexler.net/software/touchosc

[14] https://cycling74.com/products/max/

[15] https://puredata.info/

[16] http://forumnet.ircam.fr/product/openmusic-en/

[17] http://forumnet.ircam.fr/product/spat-en/

[18] http://www.mshparisnord.fr/hoalibrary/en/

[19] https://www.openal.org/

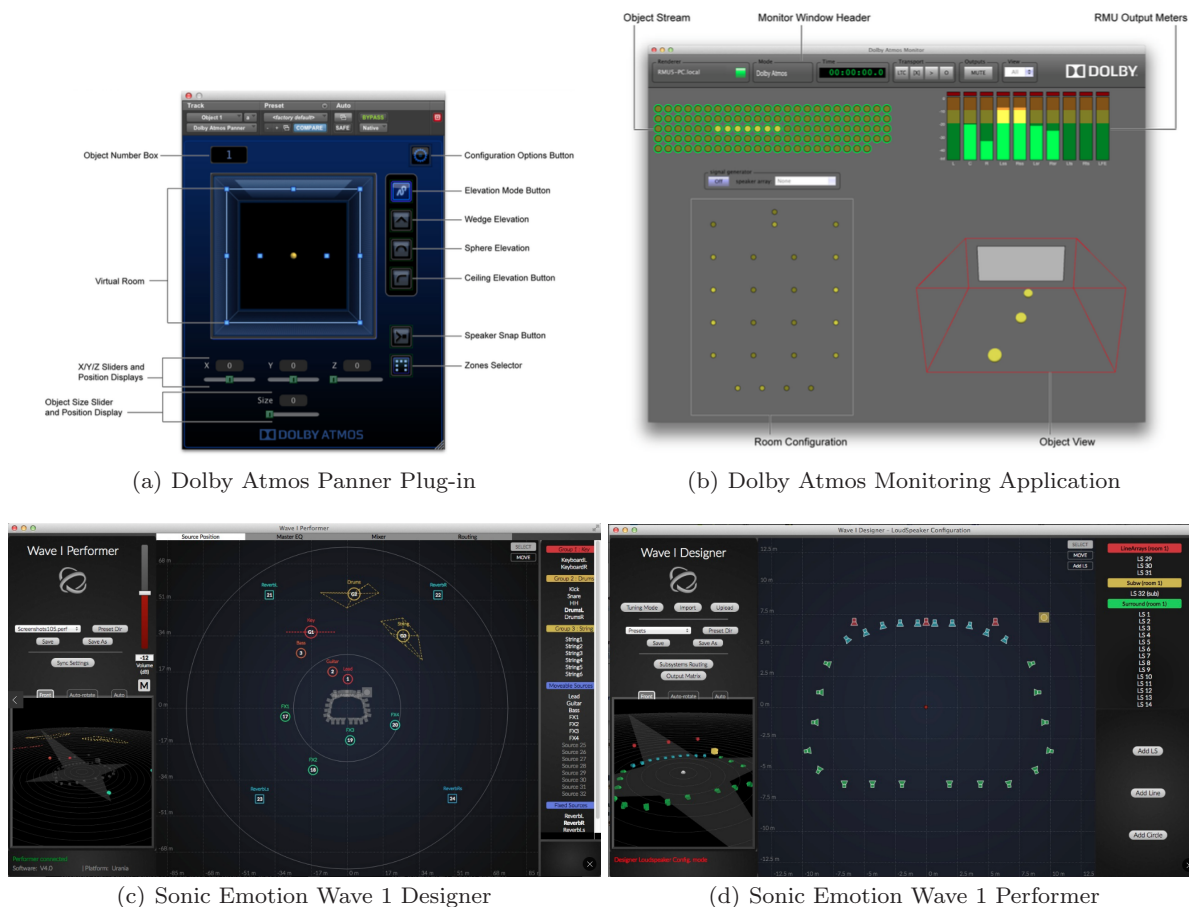[20] http://www.mathworks.com/matlabcentral/linkexchange/links/2260-spatial-audio-matlab-toolbox-v-1-0?

(a) Dolby Atmos Panner Plug-in

(b) Dolby Atmos Monitoring Application

(c) Sonic Emotion Wave 1 Designer

(d) Sonic Emotion Wave 1 Performer

Fig. 1: Examples of consumer-based 3D audio tools

3 Spatial Audio Online Survey and Questionnaire

3.1 Methodology

Our survey and questionnaire targeted spatial audio professionals (composers, recording/mixing engineers, artists, designers, and researchers) and was administered on-line. Participants were recruited by emails sent to academic institutions and on-line communities related to spatial audio. Responses were accepted for 3 months and could be edited at any time during that period. Participants were not required to answer to each question. The questions focused on the tools, techniques, and limitations of certain phases of audio production. Responses were recorded as multiple choice-single response (MC-S), multiple choice-multiple response (MC-M), Likert's scale (from 1 to 5), and free responses. All MC-M and some MC-S questions provided an "Other" option for participants to enter their own responses. Overall, questions were categorized into 5 sections:

1. *Profession and Experience* - to gather demographic information about the participants (gender, age, profession, city, country, roles, and years of experience as a practitioner).
2. *Hardware and Software Tools* - to gather a list of tools and better understand the techniques and limitations of actual authoring tools and systems.
3. *Composing/Planning and Editing* - to better understand how participants sketch, compose, and edit spatial parameters (position, orientation, directivity and reverberation) as well as limitations of these techniques.
4. *Notation and Performance* - to understand how they visually express, communicate, and perform spatial ideas.
5. *Listening/Monitoring* - to ask about their tools and methods for listening and monitoring spatialization choices, providing insight on techniques, types of visualization, and limiting factors in current systems.

The survey and questionnaire concluded with an opportunity for participants to provide additional free com-

ments and feedback. A copy of the full version of the questionnaire is available on-line[21].

3.2 Results

We collected responses from 48 participants, 47 male and 1 female, from Europe (62.5%), North America (27.1%), Australia (4.2%), and Jordan (2.1%). Ages varied between 18 and 55+ with 70.8% of them having 5+ years of experience with audio spatialization technologies. Many identified themselves to be versatile users with a variety of roles with regards to their projects. 62.5% identified themselves as Music/Audio Researchers, 58.3% as Composers, 50.0% as Recording/Mixing Engineer, and 41.7% as Sound/Installation Designer. In addition, many selected to use audio spatialization for multiple types of projects, listening formats, and virtual spaces. The majority indicated to use spatialization techniques for Recorded Music Performances (66.7%), Live Music Performances (56.3%), Sound Installations (56.3%), and Computer Aided Composition (35.4%), and they indicated to use Simple Cubic Rooms (45.8%), Speakers (39.6%), Theatres or Auditoriums (31.3%), and Spheres (29.2%) as virtual spaces for those projects (see Fig. 2). This highlights how audio spatialization techniques can be applied to a large variety of projects, and suggests the need for flexible tools to create projects with multiple listening formats and virtual spaces.

Although all participants use computers, they also use other types of hardware in their production setups, such as physical controllers (47.9%), multi-channel mixers (39.6%), multi-touch tablet (33.3%), and effects processors (29.2%). Concerning software, participants indicated DAWs (45.8%) and programming languages (textual (16.7%) and visual (27.1%)). However, MaxMSP and PureData were heavily mentioned when asked to list types of software used for 3D spatialization. In addition to those languages, many participants listed various types of plug-ins for DAWs, Ambisonic based software, and programming libraries for spatialization (see Fig. 3). Given the variety of low-level software mentioned, the high level of experience, and the technical roles of many participants, it suggests that many of them have a strong technical background in spatialization technologies. In addition, their responses and point of view helped to identify specific uses and limitations of current technologies that we discuss along four categories: *Audio Rendering*, *Visual Feedback*, *Functionality*, and *Workflow Integration*.

3.2.1 Audio Rendering

Some participants explicitly mentioned to use tools that supports Ambisonics rendering (29.2%), Stereo/Surround Panning (16.7%), VBAP (10.4%), DBAP (8.3%), Binaural (8.3%), WFS (4.2%), and ViMic (2.1%) (see Fig. 4). Participants also commented on the limitations of these renderers and how they affect the listening experience.

One participant pointed out how they use Ambisonics "*for panning as it fills holes well... but can't get sounds to come close to the listener*" and that Distance-Based Amplitude Panning (DBAP) leaves a "*bigger gap between speakers than Ambisonics*" (P_{18}). In addition, another participant mentioned that "*binaural rendering needs head tracking to accurately experience the 3D audio project*" (P_6). Others mentioned specific limitations with the design of rendering systems in that some were designed for listeners to be in the center of the configuration (sweet spot) (P_{38} & P_{45}), and that there is a need for "*easy adaptability of 'completed' work from one playback space to another*" (P_{35}). This highlights the heterogeneity of 3D audio rendering systems, which therefore requires specific mixing tools that are hardly usable from one system to the other. Addressing these technical limitations might improve the accuracy of the intended spatial rendering and provide flexibility to compose for various systems and rooms, which would in turn lead to design more flexible and generic 3D audio mixing tools adapted to the variety of systems, formats, virtual spaces and projects.

> **Limitation 1** – *The heterogeneity of 3D audio rendering systems requires specific mixing tools that limit flexibility for content production.*

3.2.2 Visual Feedback

Participants indicated that their tools provide various visual representations of position, orientation, directivity, and time. Most of their authoring tools provide a 2D visual representation (68.8%), loudness levels (54.2%), and 3D visual representation (47.9%) to represent source position or movement (see Fig. 5(a)), but many of them do not use visual representation for orientation (37.5%) and directivity (41.7%). As well, 62.5% of the participants indicated that their tools display a moving point in space to represent trajectories, but 22.9% noted that their tools do not provide any visual representation for trajectories (see Fig. 5(b)).

Even though many noted to have visualizations available in their system, they also mentioned difficulties with the available visual representation: "*deceptive spatialization tools (for example, panners which suggest*
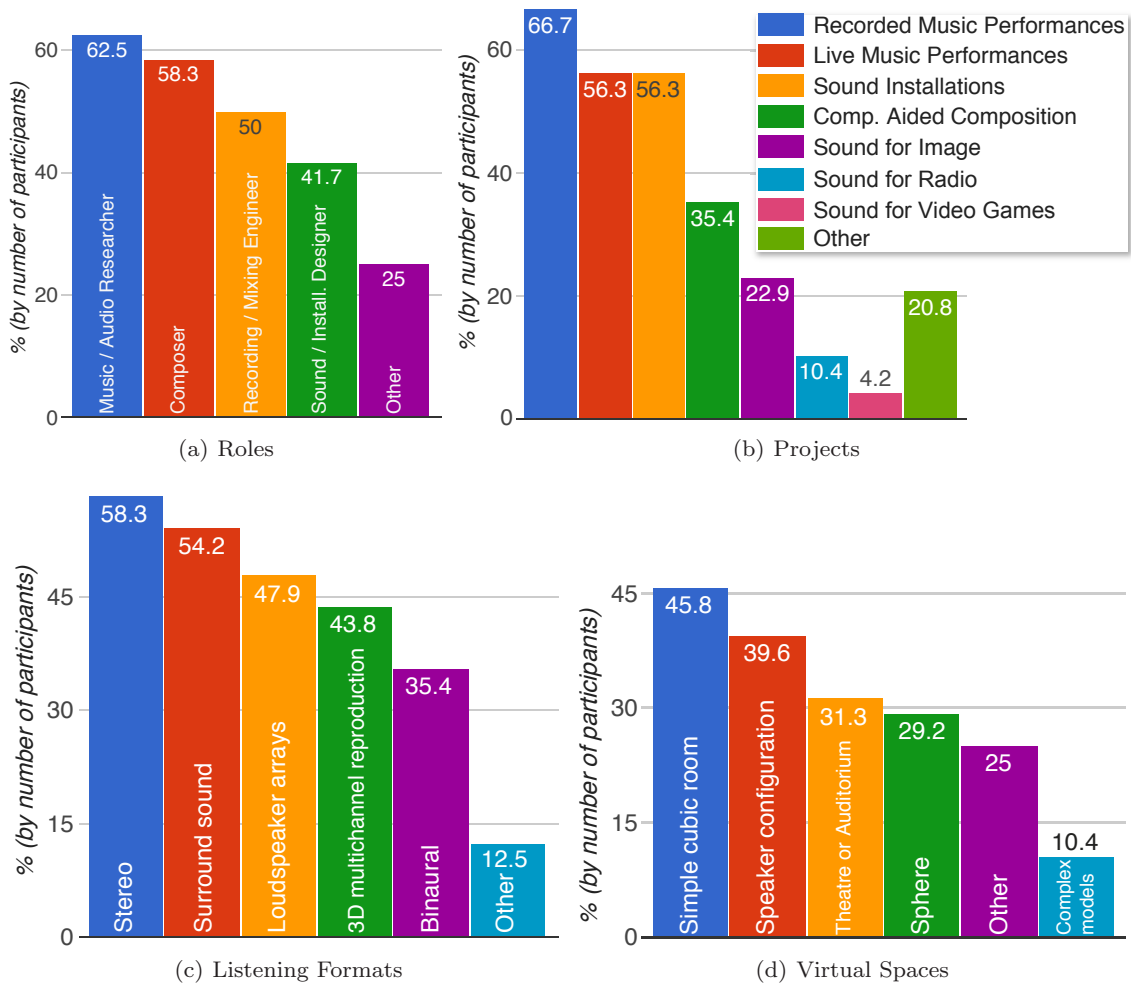
Fig. 2: The variety of *Roles* (a), *Projects* (b), *Listening Formats* (c) and *Virtual Spaces* (d) in 3D audio highlights the need for flexible spatialization tools.

Fig. 3: A word cloud visualization of the spatialization software used by participants. Text size relates to the number of participants who mentioned to use the corresponding software.
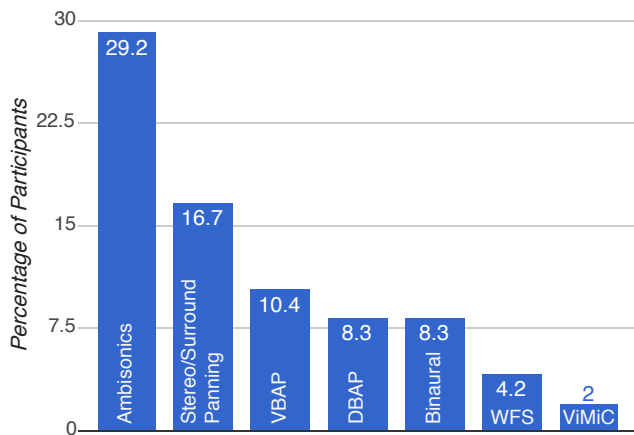
Fig. 4: The 3D audio renderers mentioned to be used by the participants in their projects.
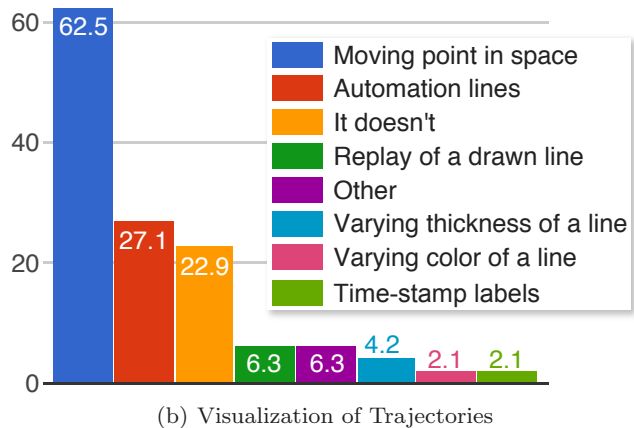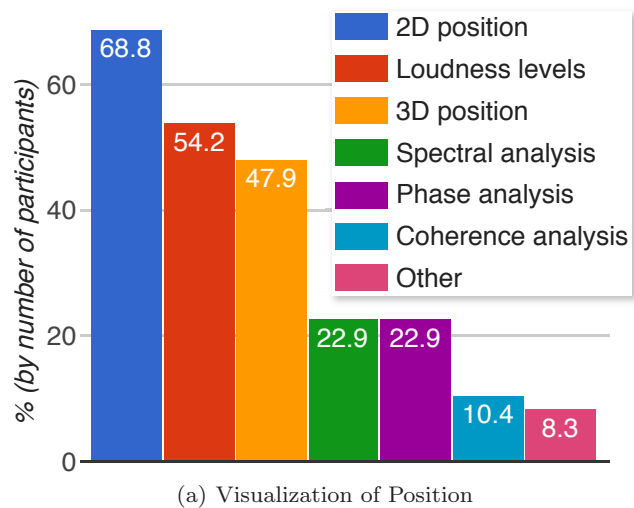

(a) Visualization of Position

(b) Visualization of Trajectories

Fig. 5: Types of visualization participants use to indicate location in a trajectory (a) and verify placement and movement (b).

sounds can be positioned in the center of an array)" (P$_{14}$);

"*many tools to monitor the spatialization in [MaxMSP,] but all these tools have advantages and drawback and I'm forced to use several tools to monitor (and sometimes re-edit) the spatialization*" (P$_{10}$). In addition, some felt that their visualizations lacked a "*direct view of the surround speakers from audience positions*" (P$_{35}$), representations of the "*position of audience members in the room relative to speaker placement*" (P$_{35}$), and "*quality of the recordings and the decoders*" (P$_{38}$). Another missing feature was better interaction with the visualization to aid comprehension, such as the ability "*to zoom in a specific area or to choose to visualize only one or a bunch of selected sources*" (P$_{38}$) and "*efficient visualization interfaces and a good gestion of time (easy zooming in a specific part of the time line)*" (P$_{26}$) for notating spatialization. Even though many of these missing features are possible solutions for improving visual feedback in spatialization tools, the need for "*a proper visualization for multichannel setups that displays what is happening in a room in a way that is easy and fast enough to read while providing sufficient detail*" (P$_8$) summarizes the main issue with current visualizations.

Limitation 2 – *The visual feedback of many authoring tools are not easily comprehensible due to inaccurate visual representations and a lack of interaction to help understand how the tool and its parameters affect the listening experience.*

3.2.3 Functionality

In addition to the issues with visual feedback, a significant amount of comments mentioned intentions and issues in producing 3D content, which have been categorized here into methods and limitations within: *Audio Recordings*, *Trajectories*, and *Input Devices*.

Audio Recordings – Fifty percent of the participants identified with the role of *Recording/Mixing Engineer* in their projects (see Fig. 2(a)). Some listed the use of proper microphone placement to capture spaciousness in their audio recordings rather than using spatialization software (see Fig. 3). In addition, participants noted the use of advanced recording techniques, associated limitations, and their effect on the listening experience. For instance, a few participants described the use of Ambisonic recordings to capture space: "*In my model, sounds are usually treated as spaces rather than as point sources. This is because I mostly work with stereo or B-format recordings of environmental sound. Occasionally a sound is defined as a point or as a directional source*" (P$_6$). In addition, P$_{35}$ mentioned the use of "*Ambisonic recording of a particular environment...*"

to create a listening experience that "*is hopefully very true to the recorded environment*". However, one participant pointed out the inability "*to adapt B-format recordings to non-standard reproduction systems for reproduction, and being able to use them as a source with richer and more detailed spatial information, while not being tied to the sweet-spot regime*" (P_8). In addition, one participant mentioned that there are "*difficulties to qualify decoders and the listening experiences*" (P_{38}). Even though recording techniques provide capabilities to capture and encode spaciousness in the audio data, there is a need to interact with their spatial characteristics and adapt them to custom speaker configurations.

Trajectories – Participants indicated various patterns and methods for inputting and editing parameters of trajectories (see Fig. 6), and their comments highlighted that there is not a consistent method in trajectory composition and edition. For predefined patterns, participants selected to use *circular path in azimuth* (64.6%), *line* (56.3%), *circular path in elevation* (33.3%), and *spiral* (25%) trajectory patterns (see Fig. 6(a)). When not using patterns, they selected to input *position and velocity at the same time* (31.3%), *automation lines not in time* (27.1%), *each positional parameter in time* (25.0%), and input *path and velocity separately* (18.8%) (see Fig. 6(b)). To edit these parameters, participants mentioned methods similar to the ones they use to initially input parameters, such as *redraw the automation lines not in time* (25.0%), *re-record automation lines for each parameter* (25.0%), *redraw path or control velocity separately* (25.0%), and *redraw the path and velocity at the same time* (22.9%) (see Fig. 6(c)).

In addition to these broad responses, participants selected to interact with the parameters both in real-time and not in real-time when producing (18.8%) and editing (16.7%) trajectories, with one noting how he "*often [makes] a recording or two, and then just adjust automation*" (P_{43}). Some difficulties for producing and editing trajectories were mentioned as well, especially for participants who only use a computer for production: the "*process to define source trajectories is very time consuming and not very intuitive*" (P_{11}); "*It can be rather time consuming [to generate] dynamic sound scenes since the changes in source properties must all be [inputted] in a somewhat unintuitive manner by hand*" (P_{36}). Others mentioned some missing functionalities, such as: "*controlling trajectories of groups of sounds with a kind of group behavior.*" (P_{44}); "*tools with more options of (controllable random position in controllable area, probabilistic control, typical trajectories such as line, circle, ellipse, square, rectangle, go back and forth, ...)*" (P_{26}). In general, our results indicate that many

tools to define and edit trajectories parameters are under developed, especially for desktop interaction.

Input Devices – Limitations in inputting and editing spatial parameters may be partially due to the input devices or control methods. Participants selected to use *software controllers* (52.1%), *physical controllers* (39.6%), *textually* (29.2%), and *drawing* (33.3%) or *pointing* (25.0%) devices to control trajectory parameters (see Fig. 6(d)). In the context of performance, the majority of participants selected that spatialization is performed through *computer software* (54.2%), *pre-defined in the software* (25.0%), or with *midi-controllers* (22.9%), but there were participants who mentioned general issues with the input devices they use. One participant, (P_{40}), felt that the typical computer mouse was limiting and that a tactile interface would offer more degrees of freedom. Others mentioned that "*distance control*" (P_{31}) and "*3D gestural controls*" (P_3) were limiting factors of their input devices. Even though more standard mixing interfaces are now including advanced input device, as mentioned previously, many participants still find limitations when interacting with input devices in general. This was summarized best by one participant who noted "*comprehensive interfaces are lacking in control of the many aspects of soundfields and the sources within them*" (P_{39}).

> **Limitation 3** – *Current interactive 3D audio tools have limited functionality for spatial manipulation of audio recordings and adaptation of various recording formats to their renderers, for authoring 3D trajectories, and the control capabilities of available input devices.*

3.2.4 Workflow Integration

Even though DAWs are widely used for spatial audio production, many participants also use programming languages and dedicated libraries (see Fig3). In addition, some comments highlighted difficulties in audio spatialization with DAWs: "*DAWs are inadequate tools. Having to arrange rendered audio to suit a specific number of output channels is tedious in the extreme*" (P_{23}); "*tools [are] still often experimental... [and] difficult to integrate with classical tools like [ProTools] or other sequencers*" (P_{26}). The complexity can become greater with larger amounts of audio due to the "*huge amount of data to control all parameters of all objects to spatialize*" (P_{47}). Even though it is difficult to incorporate 3D spatial audio tools into classic DAWs, it can be useful as an editing tool as mentioned by P_7 who
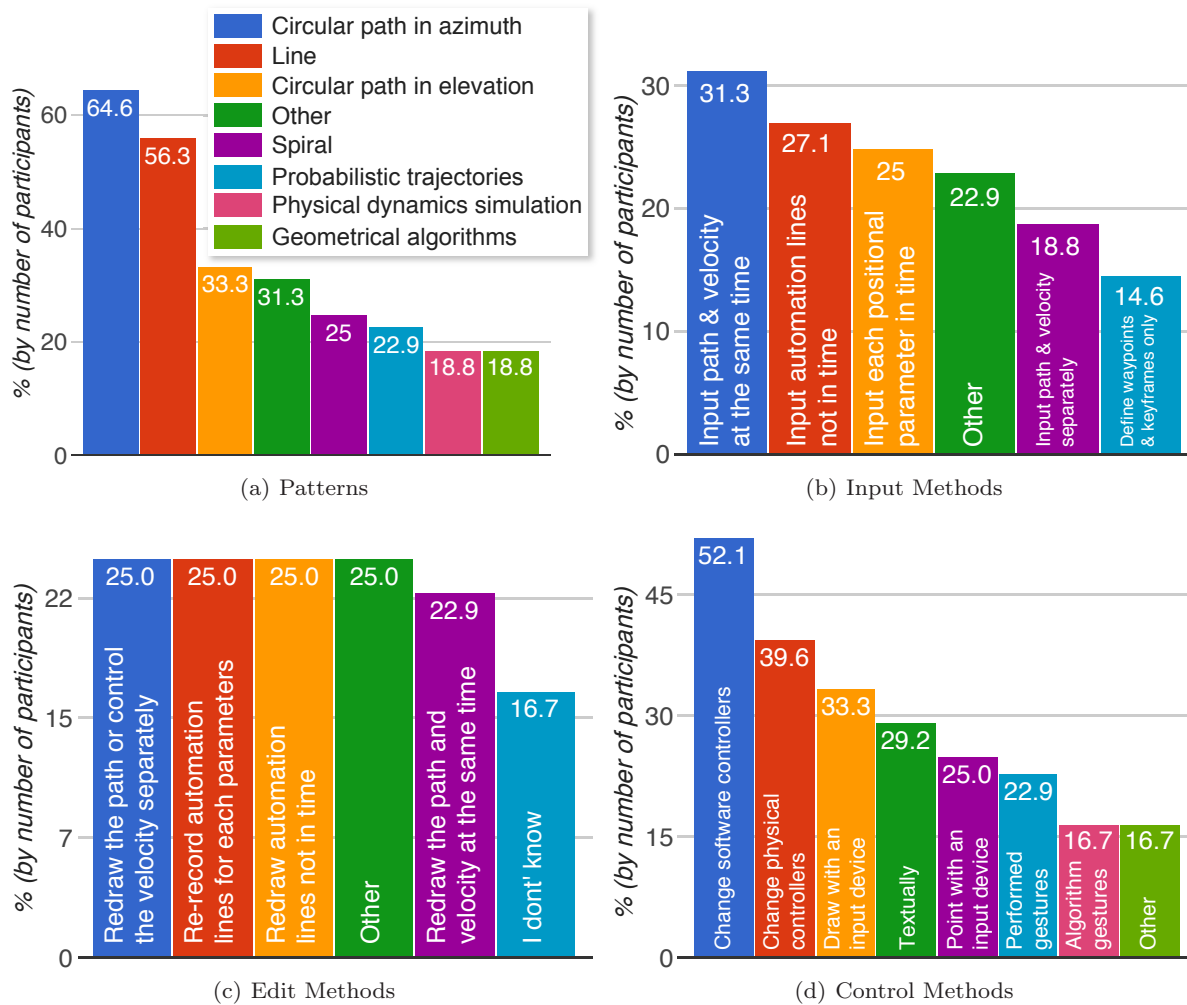
Fig. 6: The way participants edit trajectories with *Patterns* (a), *Input* (b), *Edit* (c) and *Control* methods, highlights the variety of methods to define parameters in time with existing tools.

"use[s] the HOALibrary and ICST Ambisonics externals for MaxMSP to do the actual spatializing", but uses "Reaper for any editing (and some occasional spatializing with their built-in tools)". Another mentioned that it was dependent "on the production and purpose. For real time manipulation, [MaxMSP] is mostly preferred. For precise non real time productions, [Nuendo] is a powerful tool" (P$_{18}$). There is still an attraction to use DAW software for authoring and editing 3D audio, but the design has made it inflexible with configuring 3D audio renderers and interaction. Instead, users appropriate DAWs for basic audio production tasks and use specific 3D audio libraries and programming languages for spatialization rendering, which might interrupt and limit their creative workflow.

Limitation 4 – *The lack of flexibility in common production systems to integrate advanced rendering, authoring, and editing methods tai-lored for spatialization causes 3D audio tasks to be complex and tedious, causing users to custom configure various types of tools and methods to realize their creative ideas.*

The results of our survey helped us to identify specifics needs for improving audio rendering, visual feedback, functionality, and workflow of 3D audio tools. Comments were made on the effects renderers have on listening experiences and the constraints they impose on the listeners position and playback systems. Our participants discussed that visual feedback provided by their tools have limited interaction and are difficult to comprehend, and that many available tools have limited functionalities for manipulating 3D audio recordings, composing trajectories, and controlling spatial parameters with input devices.

Lastly, participants noted problems related to the incompatibility of DAWs with 3D audio tools, which
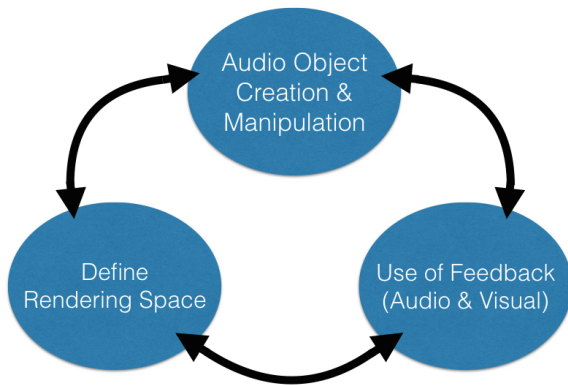
Fig. 7: Classification of three major tasks in 3D audio production workflow.

requires them to switch between multiple tools of different purposes, levels and interaction paradigms. For a more comprehensive 3D audio production systems, a higher-level classification of tasks can better organize the overall workflow and address the lower-level interaction limitations, which we identified as *Defining the rendering space*, *Creation and Manipulation of Audio Objects*, and the *Use of Feedback*.

4 Interaction for 3D Audio

In designing user interfaces for creative purposes, Terry and Mynatt (2002) identified that creative users require the ability to experiment, explore variations, and evaluate states. These needs were also in our results for the needs of 3D audio production. It has already been established in many current systems that there is a need to experiment and explore properties of audio data and spatial parameters, which can be considered as audio objects (Geier et al, 2010). The results of our survey indicate this and that there is a need for experimentation and exploration of renderers, listening formats, and reproduction spaces. In addition, participants also commented on the need for better representation within the visual feedback to comprehend the properties of rendering systems and audio objects. To better organize this in a comprehensive approach, we classify three basic tasks in 3D audio production where users experiment, explore, and evaluate: *Defining the Rendering Space*, *Creation and Manipulation of Audio Objects*, and *Use of Feedback*. This high-level classification identifies major tasks in the 3D audio workflow and helps organize how lower-level limitations can be addressed within them from a top-down approach.

4.1 Defining the Rendering Space

Defining the rendering spaces that produce/reproduce 3D audio depends mostly on the renderers themselves, but speaker properties, listener properties, and room acoustics have an effect on the outcome of the production as well. For a 3D audio loudspeaker system, users must define the renderers for use and the design of the speaker configuration. They also must consider the location of the listeners and the acoustics of the space. In some instances, such as theatres, the space may dictate the design of the speaker configuration and choice of renderers to ensure proper spatial impression can be given to the listeners. This dependency can also be reversed, like a pre-existing production that will be produced in a new space. In general, defining the rendering space requires an understanding of the dependencies between these properties that can be further explained through the subsequent four required tasks:

- *Define the renderer(s)*
- *Define listening format(s) properties*
- *Define listener(s) properties*
- *Define room acoustic properties*

The results of our survey highlight how participants experiment and explore variations with the properties of the rendering spaces. Participants used multiple types of renderers and listening formats already, but there is a lack of a homogeneous method to increase the capabilities of renderers and the flexibility to render across multiple formats. In addition, there are limited capabilities to produce for other spaces with different setups and acoustic properties without evaluation in the actual space. Improving these capabilities and functionalities can aid in the flexibility of renderers, which can also improve the capabilities of production systems and user interfaces.

4.2 Creation and Manipulation of Audio Objects

The creation and manipulation of audio objects in a 3D audio scene are the primary creative tasks (experimenting and exploring variations) in 3D audio production. In these tasks, users must first create audio objects where experimentation occurs, and which then allows them to explore variations through sonic and/or spatial manipulation of its audio and spatial capabilities.

Users create audio data through either audio recording techniques, virtual instruments, or other sonification methods and configure it to a required or chosen renderer. In turn, this and rendering capabilities define the manipulable spatial parameters of the audio object.

Manipulating the audio object consists of manipulating both audio parameters and spatial parameters in real-time or within a sequencer. As mentioned by our participants, the design of DAWs are a popular method for creating and manipulating audio data through the channel and plug-in design, but it is difficult to incorporate flexible 3D audio tools. In addition, 3D audio tools themselves have issues with regards to manipulating the spatial characteristics of audio recordings, creation and edition of trajectories, and input devices used for them. Interaction techniques have been proposed in related research, but many techniques have had mixed results from evaluation methods. Improving the capabilities and interaction techniques for the limitations mentioned can increase the power of user interfaces for 3D audio tasks.

4.3 Use of Feedback

Evaluation of past, current, and future choices are an important part of the creative process (Terry and Mynatt, 2002). As with all user interfaces, rendered feedback provides information on the states of interactive parameters, normally through visual displays. However, 3D audio production requires evaluative methods through both audio and visual feedback, which are required for the tasks for both defining the rendering spaces and audio objects.

- *Feedback of Rendering Space*
- *Feedback of Audio Objects*

Providing feedback of the rendering space not only helps define the space, but also can help better understand how their choices affect the outcome of the listening experience. Feedback of the audio objects provides a real-time evaluation method of their sonic and spatial choices, and interaction techniques with the feedback, like solo-ing or muting, can help users compare audio objects to one another. Visual feedback specifically helps users monitor the states of rendering space and audio object parameters. As mentioned by participants, current tools' visual feedback in representing the rendering space(s) and audio objects have their respective limitations, but choices are eventually decided through aural feedback in most cases. Synchronization of real-time audio/visual feedback has been mentioned as a user need before (Peters et al, 2011), but many standard DAWs only provide synchronized feedback during recording of audio data or automation parameters. In composing and editing techniques of spatial parameters, many tools only allow users to interact with visual

feedback of the GUI and evaluate aurally afterwards. Developing the methods and techniques for more comprehensible and synchronized feedback of the rendering space and audio objects can help users to evaluate in real-time of all interactions with the rendering space and audio objects.

Overall, the results of our study imply that there are high-level tasks for a general workflow for 3D audio productions across all domains. Within this classification, users go between defining a rendering space, creating and manipulating audio objects, and evaluating the creative intentions through the use of both audio and visual feedback. Developing a more comprehensive and flexible user interface requires improvements across these basic tasks, which fulfills the needs for creative user interfaces while also addressing specific needs for 3D audio production.

5 Conclusion

The design of new digital tools for creative purposes is a difficult task due to the continuous development of technologies and users constantly finding new ways to appropriate them for creative ideas. Gathering information on user's current practices, methods, intentions, and opinions helps advance capabilities to match user needs. The current study focused on gathering such information for 3D audio production through responses to an online survey highlighting specific methods and limitations in *Audio Rendering*, *Visual Feedback*, *Functionalities*, and *Workflow Integration* of current practitioners' tools and systems. Furthering this analysis, we were able to provide a classification of three basic tasks in producing 3D audio: *Defining the Rendering Space*, *Creation and Manipulation of Audio Objects*, and *Use of Feedback*. This result provides a framework which identifies general tasks within 3D audio production that covers the need for better workflow and low-level issues. Overall, the results of this study help inform the design of 3D audio production systems to include flexible rendering methods, advanced interaction techniques, and comprehensible visual feedback.

References

Boutard G, Guastavino C (2012) Archiving electroacoustic and mixed music. J of Documentation 68(6):749–771

Carrascal J, Jordà S (2011) Multitouch interface for audio mixing. In: Proc. Intl Conf on New Interfaces for Musical Expression, NIME '11

Churnside A, Pike C, Leonard M (2011) Musical movements-gesture based audio interfaces. In: Proc. Audio Eng Soc Conv 131, URL `http://www.aes.org/e-lib/browse.cfm?elib=16022`

Favory X, Garcia J, Bresson J (2015) Trajectoires: a mobile application for controlling and composing sound spatialization. In: 27ème Conf Francophone sur l'Interaction Homme-Machine., ACM, Toulouse, France, p a5, DOI 10.1145/2820619.2820624

Garcia J, Bresson J, Carpentier T (2015) Towards interactive authoring tools for composing spatialization. In: Proc. IEEE Symp on 3D User Interfaces 10, Arles, France, pp 151–152

Geier M, Ahrens J, Spors S (2010) Object-based audio reproduction and the audio scene description format. Organised Sound 15:219–227, DOI 10.1017/S1355771810000324

Gelineck S, Korsgaard D (2015) An exploratory evaluation of user interfaces for 3d audio mixing. In: Proc. Audio Eng Soc Conv 138, URL `http://www.aes.org/e-lib/browse.cfm?elib=17636`

Kendall GS (1995) A 3-d sound primer: Directional hearing and stereo reproduction. Computer Music J 19(4):23–46

Mackay W (2004) The interactive thread: Exploring methods for multi-disciplinary design. In: DIS '04: Proceedings of the 2004 Conf on Designing interactive systems, ACM, New York, NY, USA, pp 103–112

Mathew J, Huot S, Blum A (2014) A morphological analysis of audio-objects and their control methods for 3d audio. In: Proc. Intl Conf on New Interfaces for Musical Expression, Goldsmiths, University of London, London, United Kingdom, pp 415–420

Melchior F, Pike C, Brooks M, Grace S (2013) On the use of a haptic feedback device for sound source control in spatial audio systems. In: Proc. Audio Eng Soc Conv 134, URL `http://www.aes.org/e-lib/browse.cfm?elib=16742`

Peters N, Marentakis G, McAdams S (2011) Current technologies and compositional practices for spatialization: A qualitative and quantitative analysis. Computer Music J 35(1):10–27

Pulkki V (1997) Virtual sound source positioning using vector base amplitude panning. J Audio Eng Soc 45(6):456–466, URL `http://www.aes.org/e-lib/browse.cfm?elib=7853`

Terry M, Mynatt ED (2002) Recognizing creative needs in user interface design. In: Proc. 4th Conf on Creativity & Cognition, ACM, New York, NY, USA, C&C '02, pp 38–44, DOI 10.1145/581710.581718

Wessel D, Wright M (2002) Problems and prospects for intimate musical control of computers. Comput Music J 26(3):11–22, DOI 10.1162/014892602320582945