User-Defined Mappings for Spatial Sound Synthesis

Henrik von Coler
Audio Communication Group
TU Berlin
voncoler@tu-berlin.de

Steffen Lepa
Audio Communication Group
TU Berlin
steffen.lepa@tu-berlin.de

Stefan Weinzierl
Audio Communication Group
TU Berlin
stefan.weinzierl@tu-berlin.de

ABSTRACT

The presented sound synthesis system allows the individual spatialization of spectral components in real-time, using a sinusoidal modeling approach within 3-dimensional sound reproduction systems. A co-developed, dedicated haptic interface is used to jointly control spectral and spatial attributes of the sound. Within a user study, participants were asked to create an individual mapping between control parameters of the interface and rendering parameters of sound synthesis and spatialization, using a visual programming environment. Resulting mappings of all participants are evaluated, indicating the preference of single control parameters for specific tasks. In comparison with mappings intended by the development team, the results validate certain design decisions and indicate new directions.

Author Keywords

User-Defined Mapping, Spatial Audio, Spectral Modeling, Spectro-Spatial Synthesis

CCS Concepts

•**Human-centered computing** → **User studies;** •**Applied computing** → *Sound and music computing;* Media arts;

1. INTRODUCTION

A digital musical instrument (DMI) is usually understood to be comprised of a gestural controller, respectively an interface, and a sound production engine, as well as the mapping between these components [24]. Spatial sound synthesis, as introduced in the following paragraph, introduces additional requirements to the mapping problem. The *interface-mapping-synthesis* paradigm is therefor extended, since the sound reproduction system is regarded an equally important element of the instrument. Consequently, this involves the spatialization process which couples the generated sound with the reproduction system. Mapping thus needs to link the interface with sound synthesis and spatialization, resulting in the extended DMI model in Figure 1. Spatial and timbral attributes of the instrument can hence be controlled coherently. The study presented in this paper aims at exploring this interaction by letting users define mappings in spatial sound synthesis to their liking.

NIME'20, July 21-25, 2020, Royal Birmingham Conservatoire, Birmingham City University, Birmingham, United Kingdom.

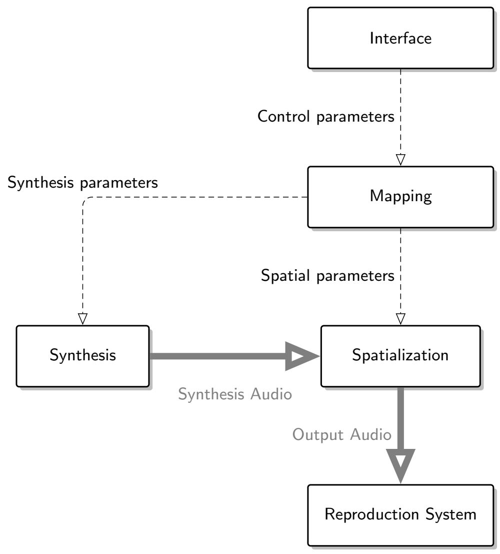


Figure 1: Extended DMI model with mapping between interface, spatialization and synthesis.

1.1 Spatial Sound Synthesis

Spatial sound synthesis refers to methods which perform an individual spatialization of single sound components at an early stage of of the synthesis process. Thereby, spacial sound structures are created, with a geometry or distribution related to spectral and temporal properties of the sound. This concept can be realized for members of all known sound synthesis families.

For granular synthesis and related approaches, like corpus-based concatenative synthesis, spatialization is realized in the time domain, by assigning individual spatial attributes to single grains, or units [16, 6]. Spatial behaviour of units can then be linked to their audio features and controlled through gesture or algorithms. In physical modeling, *synthesis-driven wave field reproduction* has been proposed [17]. It combines physical models of strings with wave field synthesis (WFS) rendering. WFS loudspeaker driving functions are calculated by a virtual piston model with virtual physical extent and orientation. Regarding abstract algorithms, spatialization can be realized by means of FM synthesis [15]. Frequency modulation is generated though movements of virtual sound sources with the appropriate Doppler shift. In the same way, amplitude modulations can be achieved through modulations in distance. Finally, additive synthesis and spectral modeling allow the individual spatialization of spectral components, more specifically of single sinusoids and noise [21, 20, 3].

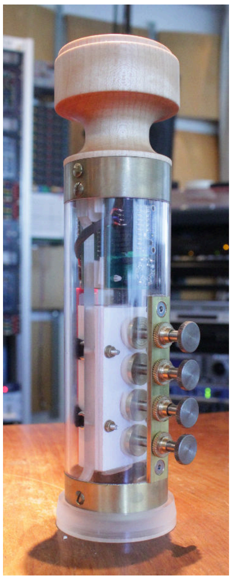


Figure 2: Front view of the BinBong MKII.

1.2 Control and Mapping

Novel interfaces designed for the spatialization of sound and music have been invented since the beginnings of electroacoustic music. Early examples include Stockhausen's rotating table for creating movements in fixed media compositions and Pierre Schaeffer's *Pupitre d'Espace* for the real-time gestural control of diffusion in acousmatic music [19].

Since then, many means for gestural control have been applied as interfaces for interacting with spatial sound. A system based on a multi-touch display for moving virtual sound sources was presented by Johnson et al. [9]. Motion capturing based on camera systems has been used for gestural control of sound spatialization in various projects [1]. Using bend sensors on the fingers and a 6 degree-of-freedom position tracker, Marshall et al. [14] created gloves for the gestural control of spatialization. Turntable-based spatialization has been proposed through visual hand tracking [13].

In addition to the design of novel interfaces, mapping between control data and sound generation has gained attention as a field of research in the past decades. The concept of user-defined mappings allows musicians to dynamically configure this aspect to their individual needs. In addition, developers and researchers gain insight into mapping strategies for improving human computer interaction in musical instrument design. A solution for user-defined mapping has been presented within the Max/MSP environment[2, 22]. *libmapper* [12] is a cross-platform software library, specifically designed for dynamic mapping between parts of interactive systems. In a related study, Brown et al. [4] evaluated user-defined mappings for data-gloves in a musical application. Results obtained through qualitative analysis indicate differences in mapping strategies between experienced and inexperienced users.

2. SYSTEM DESCRIPTION

The system introduced in this section is considered a holistic musical instrument, including a haptic control interface, an analysis-synthesis system for sound generation, as well as the loudspeaker system and rendering software used for spatialization. All relevant components will be introduced in this section, focusing on the concluding study and the aspect of mapping.

2.1 Haptic Interface

Control over synthesis and spatialization is granted through the BinBong. The haptic interface is the second version of an ongoing collaboration between the Audio Communication Group at TU Berlin and the Federal Institute for Musicology (SIM) in Berlin. Isolated aspects of the first prototype, more precisely reaction times and error rates in simple tasks, have been evaluated previously in a usability study [23]. Developers experience, user feedback and results of the study led to the development of the BinBong MKII, shown in Figure 2. The cylindrical device, equipped with wireless communication and various sensors, is designed to be operated with two hands. Main intended application of the interface is the expressive control of monophonic melody instruments, considering the ability of controlling spatial aspects.

All 9 parameters generated by the device, referred to as *control parameters*, will be introduced in this section. Several force-sensing resistors (FSRs) are integrated in the interface, grouped in three units, as shown in Figure 3, completed by an inertial measurement unit (IMU), capturing the absolute orientation in space.

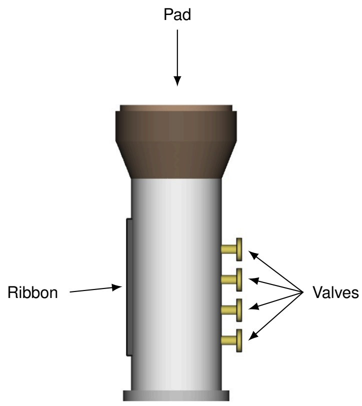


Figure 3: Force-sensing units.

2.1.1 Valves

Four mechanisms, referred to as *valves*, are organized in a row, allowing the operation with four fingers, with one FSR under each valve. Silicon cushions of the first version have been replaced with felt, which is common in instrument building, due to its durability. The mechanisms have a soft action point and can be compressed.

The valves offer two control parameters, listed in Table 1. The binary combination detects whether single valves are pressed and generates an integer parameter with $2^4 = 16$ steps [23]. This control parameter is designed to influence the pitch of the instrument within one octave. Additionally, the mean overall force applied to the valves is used as a control parameter. The overall force is scaled by the number of pressed valves, allowing the use of the full scale for any combination. Initially, this parameter was intended to control the pitch deviation within one semitone.

2.1.2 Pad

Four FSRs are located under the wooden pad on top of the cylinder, also cushioned with felt inlays to allow a soft action point. Two control parameters, listed in Table 2, are generated by the pad. The overall applied force is intended to

Table 1: Control parameters of the valves.

Parameter	Description	Range
Binary	integer value, based on binary combinations of pressed/unpressed	0 ... 12
Force	the overall force applied to the valves	0 ...1

Table 2: Control parameters of the pad.

Parameter	Description	Range
Force	overall applied force	0 ... 1
Position	shift towards front/back	0 ...1

be used for the excitation, respectively the intensity of the sound synthesis. The position parameter indicates whether force is shifted towards the side of the valves (front) or towards the ribbon side (back).

2.1.3 Ribbon

The first version of the BinBong featured three piezo - electric buttons, opposing the valves, intended to control the octave offset of the pitch. In the MK II they are replaced with a force-sensing linear potentiometer (FSLP), referred to as *ribbon*. With a length of 10 cm, the ribbon provides three control parameters, as listed in Table 3, including the applied force, the position of contact and an additional swipe gesture, which is designed to control the octave offset.

Table 3: Control parameters of the ribbon.

Parameter	Description	Range
Force	overall force applied	0 ... 1
Position	position of contact	0 ... 1
Swipe	swiping up or down increases or decreases an integer-based parameter	-5 ... +5

2.1.4 Orientation

The interface is equipped with a 9 degrees of freedom inertial measurement unit (IMU), allowing to track its absolute orientation in space. Euler angles are used to capture the rotation of the interface around three axis, as shown in Figure 4. This set of parameters, listed in Table 4, allows the use as a pointing device, intended to control the input direction of the virtual sound source.

2.2 Sound Synthesis

Sound synthesis is carried out in a standalone application, implemented as a Jack[1] client on a Linux audio system [7]. The underlying method for sound synthesis is based on a statistical model of spectral modeling data, gathered from a library of violin recordings [8]. The synthesis engine provides 24 outputs, each representing a Bark frequency band. Spectral components of the synthesis are routed to the outputs, weighted by a band-pass filter for each Bark frequency band, resulting in 24 spectrally separated signals.

Control over the synthesis engine is granted through the OSC protocol. Main parameters for controlling the synthesis engine are *pitch* and *intensity*. The full set of parameters available in the user study is listed in Table 5.

[1] `https://jackaudio.org/`

Table 4: Euler angles of the orientation sensor unit.

Parameter	Description	Range
Pitch	pointing up or down	±90°
Roll	tilted left or right	±90°
Yaw	rotated clockwise/counterclockwise	±180°

Table 5: Rendering parameters of the synthesis engine.

Parameter	Description	Range
Pitch Offset	deviation in pitch	± one semitone
Octave Offset	deviation in pitch	± 5 octaves
Intensity	intensity of excitation	0 ... 1
Tonal Level	level of the tonal part of the sound	0 ... 1
Noise Level	level of the noise part of the sound	0 ... 1

2.3 Spatialization

A hemispherical loudspeaker system with 21 Genelec 8020C loudspeakers and two subwoofers, arranged within a surface area of $\approx 6.5\,\mathrm{m} \times 5\,\mathrm{m}$ and a height of $\approx 3.5\,\mathrm{m}$ is used for sound reproduction. The Linux standalone version of Panoramix [5] is used for spatialization of the signals generated by the sound synthesis engine, using a 5^{th} order Ambisonics renderer. Each of the 24 Bark-band signals is routed to an individual point sound source.

All 24 point sources are grouped as a virtual sound source with spatial extent, as visualized in Figure 5. Similar models are used for the description of spatial sound in the domain of electroacoustic music [10] and virtual acoustics [11]. The angles azimuth (A) and elevation (E), the source distance (D), as well as the spread (S) are used as metaparameters, listed in Table 6.

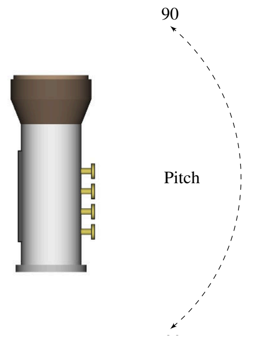
Table 6: Rendering parameters of the spatialization.

Parameter	Description	Range
Azimuth	horizontal angle of incidence	±180°
Elevation	vertical angle of incidence	±90°
Distance	distance to the source	1.3 m ... 10 m
Spread	spatial extent	≈ 0.01 m ... 5 m

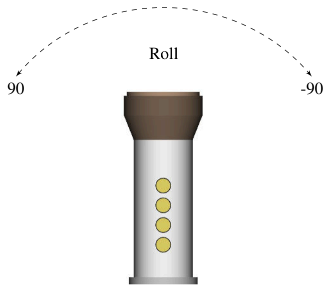
2.4 Mapping

Mapping between the *control parameters* and the *rendering parameters* is performed in Puredata [18]. During development and testing, this allows a flexible way of connecting interface and synthesis by making changes during run-time.

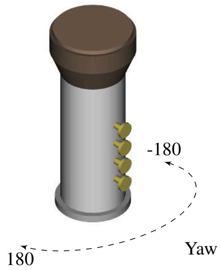
For the study presented in the following section, a simplified patch with limited options is used. Figure 6 shows a complete patch created by a participant. Control parameters are located in the upper half of the patch, grouped by control unit in colored boxes. The instantaneous value for each parameter is visualized by an individual vertical slider. Rendering parameters are located in the lower half of the patch, divided into synthesis and spatial parameters. Vertical sliders show the instantaneous value of each rendering parameter. Connections between control and rendering parameters can be established through a drag and drop action and deleted, if desired. The polarity of each rendering parameter can be inverted using a toggle switch.



(a) Pitch



(b) Roll



(c) Yaw

Figure 4: Three angles of spatial orientation.

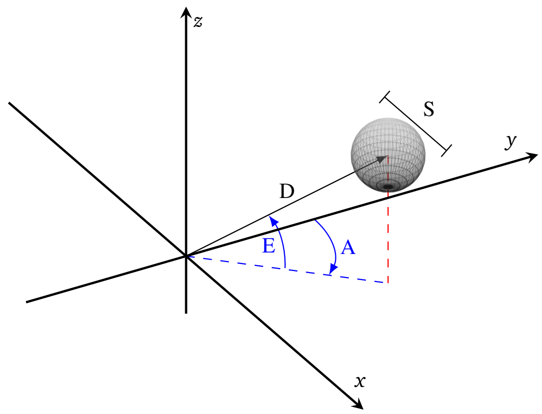


Figure 5: The virtual sound source with controllable parameters.

3. USER STUDY

The procedure proposed in this study can be considered an evaluation of the instrument and mapping procedure in itself, as well as an approach towards user-driven development. The full study consists of two parts: the *mapping stage*, during which participants create individual mappings for the system and the *gesture part*, in which participants are asked to perform certain gestures, described by written instructions. In the scope of this paper, the mapping part, respectively the individual mappings will be presented and discussed, whereas the gesture part, aiming at a subjective evaluation of the individually defined mappings, will be treated in later analyses.

3.1 Procedure

Prior to the mapping part, participants were introduced to all relevant components of the system. During the mapping stage, each participant was given 30 minutes to create his or her own mapping, with the following instruction:

> *The objective of this part is to create an enjoyable mapping, which offers the most expressive control over all synthesis and spatialization parameters.*

Four guidelines had to be taken into account when creating the mapping, allowing one-to-many mappings and excluding many-to-one connections:

1. *Every rendering parameter of synthesis and spatialization must be influenced through the mapping.*
2. *Control parameters may remain unconnected.*
3. *A single control parameter may be mapped to multiple synthesizer or spatialization parameters.*
4. *A synthesis or spatialization parameter must not have more than one control parameter connected to its input.*

During the experiment, participants could move within a marked area of about 1 m^2. Mouse and screen, used by the participants for mapping, were located at the boundary of this area. The screen was mirrored to the control room to ensure correct operation. No clock was shown to the participant when performing the task. The test management informed the participant when the 30 minutes were elapsed, leaving additional 5 minutes for finishing the mapping. A talk back microphone could be used by the participant to inform the test management when the mapping was finished earlier. The state of all connections was recorded over time. This allows the evaluation of the final mappings, as well as the analysis of the mapping process.

3.2 Sample

20 participants attended the study. Due to incomplete data, two sets are excluded from the evaluation. The remaining 18 participants have an average age of 28.5 years, with an interquartile range of 8 years. 15 male and 3 female persons took part, among them 17 right-handed and one left-handed user. German was the native language for 12 participants, Greek for two, with one Chinese, one Hungarian, one Spanish and one Italian speaker. The study was conducted in English.

Participants were asked whether they had experience with spatial audio in general and with electronic music, through a 7-point likert skale, ranging from *Completely Disagree* (0) to *Completely Agree* (6). Mean experience with spatial audio was 4.72, representing a *strongly agree*, with a standard deviation of 1.59. The mean experience with electronic music was 4.72, also a *strongly agree*, with a standard deviation of 1.28. Ten participants were master's students, three researchers at Audio Communication Group. Seven participants were professionals in research or production related to spatial audio technologies.

3.3 Results

The overall number of connections set by all participants for each control parameter is shown in Figure 7. With a total of six connections made, *Valve: Binary* is the least used control parameter. *Pad: Force*, *Ribbon: Position* and *IMU: Pitch* have been connected 20 times. A X^2-test shows no significant divergence from equal distribution (p = *0.323*).

Figure 8 shows the mapping matrix with the absolute frequencies of connections made between control and rendering parameters by all participants. Since one-to-many

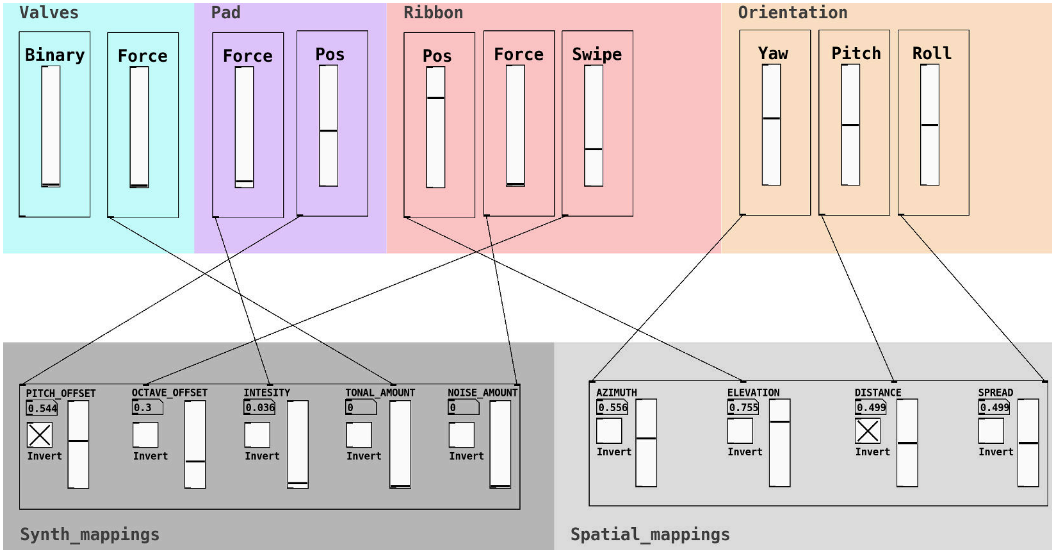


Figure 6: Mapping GUI in Puredata, as used by participants.

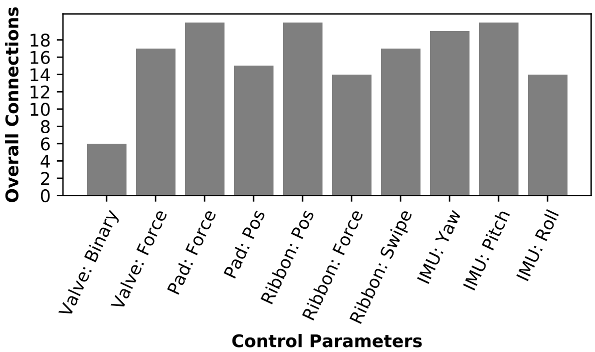


Figure 7: Frequency of use for all control parameters.

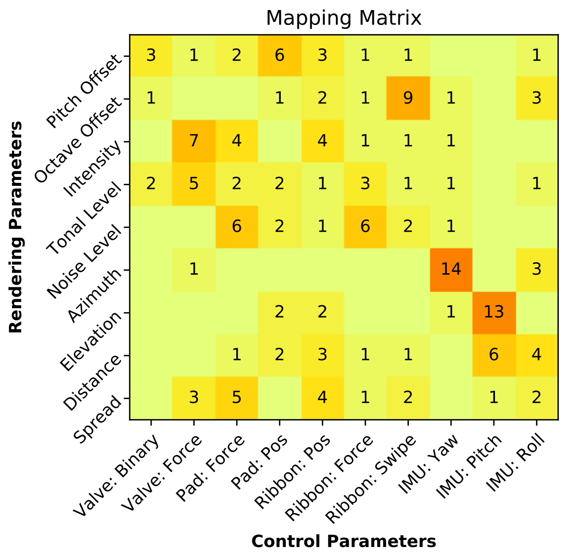


Figure 8: Mapping frequency between control and rendering parameters.

mappings are allowed and many-to-one mappings are excluded, the sum of each row is 18, whereas the sum of each column depends on the use frequency of the relevant control parameter.

Additional X^2 tests were calculated for cross-tabulations of each control parameter with the frequency of rendering parameters chosen by the participants. Results are shown in Table 7. With exception of *Valve: Binary*, *Ribbon: Position* and *IMU: Roll*, the distribution of chosen rendering parameters significantly diverges from uniform distribution for every single control parameter.

Most prominent are the mappings between the control parameter *IMU: Yaw* and the spatialization parameter *Azimuth* (14/19) as well as between *IMU: Pitch* and *Elevation* (13/20). The control parameter *Ribbon: Swipe* is connected to the synthesis parameter Octave Offset 9 out of 17 times. The control parameter *Valve: Force*, was most frequently mapped to the *Intensity* (7/19), followed by the *Tonal Level* (5/19). *Pad: Position* was most frequently mapped to the *Pitch Offeset* (6/15). *Ribbon: Pos* was often used, yet not preferred for specific rendering parameters.

3.4 Discussion

As the mapping matrix shows, certain combinations between control and rendering parameters are preferred by the participants. The predominant connections between *IMU: Yaw*, *IMU: Pitch* and the spatialization parameters coincide with the developers' intentions. Participants did not find a common use for the remaining orientation parameter *IMU: Roll*. Presumably the third orientation parameter introduces too much complexity and interacts with *IMU: Yaw* and *IMU: Pitch*. The frequent use of *Valve: Force* for the *Intensity* and the related *Tonal Level* is more operable than for the initially intended *Pitch Offset*, since changes in fingering are thus decoupled from pitch modulations.

4. CONCLUSION

The presented approach towards user-defined mappings can be considered an evaluation, as well as part of a user-driven design. Participants appreciated the task, whereby a playful exploration of the system becomes possible. Results of the mapping stage give insight on the interplay between interface and spatial sound synthesis and indicate that certain control parameters are preferably mapped to specific rendering parameters. In some cases mappings are chosen as expected, whereas other connections preferred by the users

Valves		Pad		Ribbon			IMU		
Binary	Force	Force	Pos	Pos	Force	Swipe	Yaw	Pitch	Roll
0.0591	< 0.01	0.0165	0.0323	0.5366	0.0202	< 0.01	< 0.01	< 0.01	0.1644

Table 7: X^2 test p-values for uniform distribution of chosen rendering parameters with respect to each control parameter

differ from the developers' intentions. This feedback is a valuable step for the further development of the system.

Evaluation of the gesture part and the comprehensive surveys will be part of future work. The user experience related to specific mapping strategies can then be estimated. Beyond the scope of development and testing, the system does benefit from a flexible, user-definable mapping environment when being used by performers. For this purpose, the implementation of a more convenient GUI would be necessary, featuring additional mapping options. This includes additions, scaling, offsets and non-linear amplifications.

5. REFERENCES

[1] A. Aska and M. Ritter. Approaches to Real Time Ambisonic Spatialization and Sound Diffusion using Motion Capture. In *Proceedings of the International Computer Music Conference*. Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 2016.

[2] Bevilacqua, Frédéric and Müller, Remy and Schnell, Norbert. MnM: a Max/MSP mapping toolbox. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 85–88, 05 2005.

[3] J. Bresson. Spatial Structures Programming for Music. In *Proceedings of the Spatial Computing Workshop (SCW) – Co-located w. Autonomous Agents and Multi Agent Systems (AAMAS)*, Valencia, Spain, 2012.

[4] D. Brown, C. Nash, and T. Mitchell. Understanding user-defined mapping design in mid-air musical performance. In *Proceedings of the 5th International Conference on Movement and Computing*, page 27. ACM, 2018.

[5] T. Carpentier. Panoramix: 3d mixing and post-production workstation. In *Proceedings of the International Computer Music Conference (ICMC)*, 2016.

[6] A. Einbond and D. Schwarz. Spatializing Timbre with Corpus-Based Concatenative Synthesis. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 1–1, New York, United States, June 2010. cote interne IRCAM: Einbond10a.

[7] Henrik von Coler. A Jack-based Application for Spectro-Spatial Additive Synthesis. In *Proceedings of the 17th Linux Audio Conference (LAC-19)*, Stanford University, USA, 2019.

[8] Henrik von Coler. Statistical Sinusoidal Modeling for Expressive Sound Synthesis. In *Proceedings of the 22nd International Conference on Digital Audio Effects (DAFx)*, Birmingham, UK, 2019.

[9] B. Johnson and A. Kapur. Multi-Touch Interfaces for Phantom Source Positioning in Live Sound Diffusion. In *Proceedings of the 2013 International Conference on New Interfaces for Musical Expression*, pages 213–216, Daejeon + Seoul, Korea, 2013.

[10] G. S. Kendall. Spatial perception and cognition in multichannel audio for electroacoustic music. *Organised Sound*, 15(3):228–238, 2010.

[11] A. Lindau. *Spatial Audio Quality Inventory (SAQI). Test Manual. v1.2.* TU Berlin, 2015.

[12] J. Malloch, S. Sinclair, and M. M. Wanderley. Libmapper (a library for connecting things). In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 3087–3090. 2013.

[13] G. Marentakis, N. Peters, and S. McAdams. Dj SPAT: spatialized Interactions for DJs. In *Proceedings of the International Computer Music Conference - ICMC*, volume II, pages 89 – 94, San Francisco, USA, 2007.

[14] M. Marshall, J. Malloch, and M. Wanderley. Gesture Control of Sound Spatialization for Live Musical Performance. In *Proceedings of the 7th International Gesture Workshop - Gesture-Based Human-Computer Interaction and Simulation*, volume 5085, pages 227–238, Lisbon, Portugal, 05 2007.

[15] R. McGee. Spatial modulation synthesis. In *Proceedings of the International Computer Music Conference (ICMC)*, 2015.

[16] A. McLeran, C. Roads, B. L. Sturm, and J. J. Shynk. Granular sound spatialization using dictionary-based methods. In *Proceedings of the 5th Sound and Music Computing Conference, Berlin, Germany*, number 1, 2008.

[17] A. Müller and R. Rabenstein. Physical modeling for spatial sound synthesis. In *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, 2009.

[18] M. S. Puckette. Pure Data. In *Proceedings of the International Computer Music Conference (ICMC)*, Thessaloniki, Greece, 1997.

[19] Pysiewicz, Andreas and Weinzierl, Stefan. Instruments for spatial sound control in real time music performances. a review. In *Musical Instruments in the 21st Century*, pages 273–296. Springer, 2017.

[20] M. Schumacher and J. Bresson. Spatial Sound Synthesis in Computer-Aided Composition. *Organised Sound*, 15(3):271–289, 2010.

[21] D. Topper, M. Burtner, and S. Serafin. Spatio-operational spectral (sos) synthesis. In *Proceedings of the International Computer Music Conference (ICMC)*, Singapore, 2003.

[22] Van Nort, Doug and Wanderley, Marcelo. The LoM Mapping Toolbox for Max/MSP/Jitter. In *Proceeedings of the International Computer Music Conference (ICMC)*, 2006.

[23] von Coler, Henrik, Treindl, Gabriel, Egermann, Hauke, and Weinzierl, Stefan. Development and Evaluation of an Interface with Four-Finger Pitch Selection. In *Proceedings of the 142nd Audio Engineering Society Convention*. Audio Engineering Society, 2017.

[24] M. Wanderley and P. Depalle. Gestural control of sound synthesis. *Proceedings of the IEEE*, 92(4):632–644, Apr 2004.