Sound spatialization across disciplines using virtual microphone control (ViMiC)

Nils Peters [1#], Jonas Braasch [2#] and Stephen McAdams [3#]

[1] Center for New Music and Audio Technologies (CNMAT), UC Berkeley
[2] School of Architecture, Rensselaer Polytechnic Institute (RPI), Troy
[3] Schulich School of Music, McGill University, Montreal
[#] Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Montreal

Background in Spatial Sound Perception and Synthesis. Spatial sound perception is an important process in how we experience sounds in our environment. This process is studied in the fields of otology, audiology, psychology, neuroscience and acoustical engineering. Its practical implications are notably found in communications, architectural acoustics, urban planning, film, media art and music. The synthesis of spatial sound properties by means of computer and loudspeaker technology is an ongoing research topic.

Background in History of Spatial Music. In spatial music, perceptual effects of spatial sound segregation, fusion and divided attention are explored. The compositional use of these properties dates back to the 16th century and the music by Willaert and Gabrieli for spatially separated instruments and choirs. Electroacoustic inventions in the 19th and 20th century, such as microphones and loudspeakers, and the recent increase in computer resources have created new possibilities and challenges for composers.

Aims. The aim of this project was to develop a perceptually convincing spatialization system that is flexible and easy to use for musical applications.

Main contribution. Using an interdisciplinary approach, the Virtual Microphone Control systen (ViMiC) was developed and refined to be a flexible spatialization system based on the concept of virtual microphones. The software was tested in multiple real-world user scenarios ranging from concert performances and sound installations to movie production and applications in education and medical research.

Implications. Our interdisciplinary development approach can guide other development efforts for creating user-friendly computer music tools. Due to its specific feature set, ViMiC has become a flexible tool for spatial sound rendering that can be used in a variety of scenarios and disciplines. We hope that ViMiC will motivate further creative and scientific interest in sound spatialization.

Keywords: Spatialization, Sound Perception, Human-Computer Interaction, Spatial Music.

1 Introduction

Spatialization, the synthesis of spatial properties of sounds and rooms for a listener, is a field of interest for musicians, media artists, sound engineers, audiophiles, and composers.

In the particular case of composers, many have long desired to integrate spatial dimensions into their music in addition to the traditional concepts of pitch, timbre, and rhythm. The effects of spatial sound segregation and fusion were introduced through static placement and separation of musicians in the concert space, for example, in the antiphonal music of Gabrieli (1557–1612), and later enhanced through dynamic modifications of the sound source position, e.g., by Charles Ives' father George (Zvonar, 1999). The term Spatial Music describes pieces that build on spatiality of sound as an essential part of the composition. Galeyev (2006) defines "Spatial music [...] a term referring to the experiments with musical sounds moving in space". In connection to the Pulitzer price winning composer Henry Brant, Spatial Music is defined as a genre, where the fixed deployment of musicians throughout the concert hall is part of the composition (Board, 2002).

Due to the invention and integration of microphones, tape recorders and loudspeakers, sound spatialization regained popularity for musical applications in the early 20th century. Since Disney's Fantasia (1940), it was also adopted by the motion picture industry (Klapholz, 1991). Subsequently, computer technology enabled the development and refinement of spatialization concepts such as Ambisonics (Gerzon, 1973), Wave Field Synthesis (WFS, Berkhout et al., 1993), and Vector Base Amplitude Panning (VBAP, Pulkki, 1997). Today, the real-time interaction with spatial sounds via control devices and gestures has become a subject of considerable interest (e.g., Marshall et al., 2006).

In creating new music software tools, we find that the development process inside the researcher's lab environment is often insufficient because real-world conditions and unexpected user challenges have to be taken into account. Therefore the Virtual Microphone Control (ViMiC) System was developed with an interdisciplinary approach, including studies of real-world applications for spatialization.

2 Basic ViMiC Concept

ViMiC is a tool for real-time spatialization synthesis, particularly for concert situations and site-specific immersive installations with larger or non-centralized audiences. This section summarizes the ViMiC rendering concept. Further technical details can be found in Braasch et al. (2008) and Peters et al. (2008).

ViMiC builds upon the principles of human sound localization. The human auditory system primarily utilizes interaural level and time differences to determine whether the sound is coming from the left or right. Good summaries of human sound localization are given in Blauert (1997) and Moore (2012), which also provide insight into the phenomenon of summing localizations, a perceptual effect that allows us to

pan signals between two loudspeakers to be perceived in between both of them. ViMiC produces, depending on the position of the virtual sound sources, Inter-Channel Level Differences (ICLDs) and Inter-Channel Time Differences (ICTDs) between the loudspeaker signals, which are then transformed into interaural level and time differences on the pathways between the loudspeakers and the listeners' eardrums (Braasch, 2005).

Common spatialization algorithms are based on panning laws and Inter-Channel Level Differences, not always capable of producing the spatial sound qualities a Tonmeister learned to capture by carefully arranging microphones and musicians (Borwick, 1973). ViMiC is based on these Tonmeister procedures to create a virtual sound recording scene, which consists of three main components: sound sources, a recording room, and microphones. The algorithm simulates these three components in a real-time environment. In ViMiC, sound sources are defined through their location, radiation pattern, orientation, and sound pressure level. Similarly, virtual microphones are characterized through microphone directivity pattern, recording sensitivity, position and orientation in the recording room. The parameters of the virtual recording room (size, room geometry, and surface properties) control the contribution of early room reflections and reverberation to the spatial auditory image. An overview of the system's architecture is shown in Figure 1 and is further explained in the following subsections.
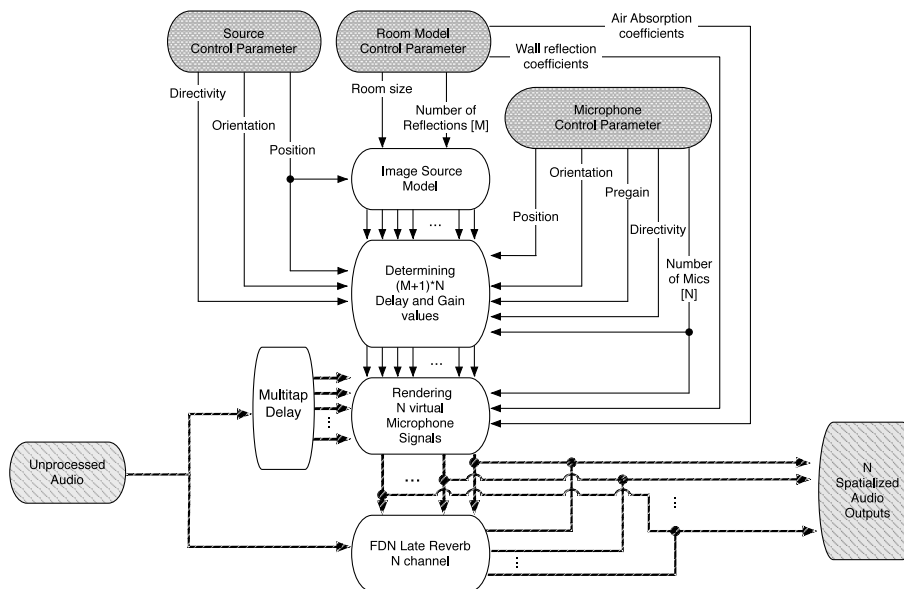

Figure 1. ViMiC architecture.

Virtual microphones have been previously employed in room acoustics simulators such as CATT to auralize the virtual room impression. While these applications aim

to simulate the physical properties of the room acoustics as accurately as possible, which results in long computation times, ViMiC is optimized to trade accuracy for real-time capabilities. The ViMiC concept should not be mistaken with the similarly named application Visual Virtual Microphone (VVMic). While VVMic is an Ambisonics tool (see e.g., Malham and Myatt, 1995) to decode loudspeaker feeds from B-format recordings based on arrangable, but always coincident microphone setups, the ViMiC system synthesizes spatial audio content in a virtual auditory environment where the microphone placement is unrestricted, as explained below.

2.1 Source – Microphone Relation

Within the virtual recording room, sound sources and microphones can be placed and moved in 3D as desired. Figure 2 shows an example of one sound source recorded with three virtual microphones. Based on the physical principles of sound propagation and the distance between a virtual sound source and each virtual microphone, the ViMiC algorithm computes the intensity and the time-of-arrival at each virtual microphone, thus creating Inter-Channel Level and Inter-Channel Time Differences.
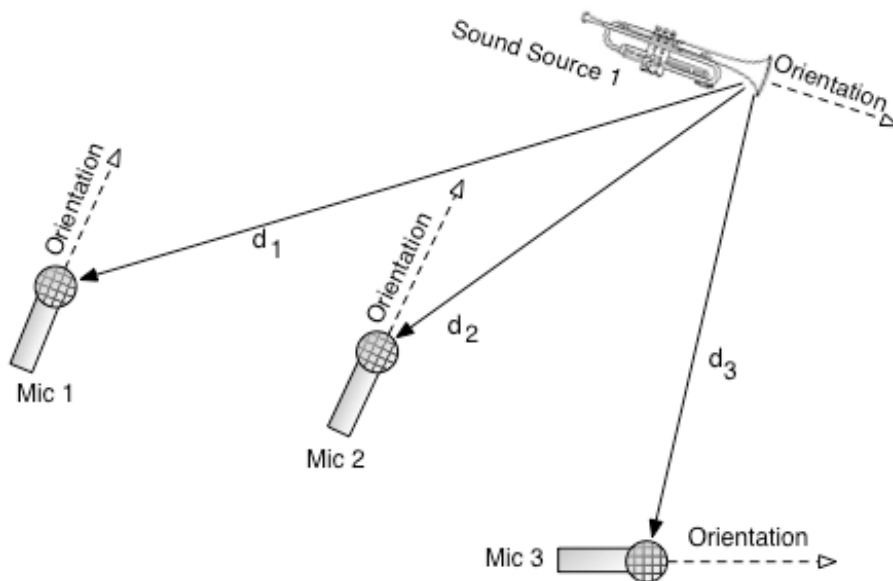

Figure 2. Source-to-microphone relation in ViMiC.

In ViMiC, the microphones have a controllable directivity. Figure 3 shows common directivity patterns found in real microphones, ranging from omnidirectional (Figure 3.a) through Cardioid patterns to Figure-8 microphones (Figure 3.d) with a negative phase component. The directivity Γ of these microphone pattern can be synthesized

with Equation 1 with the given angle of incidence δ and the coefficient α ranges from 0 (non-directional) to 1 (Figure-8 directivity):

$$\Gamma_\delta = 1 - a(1 - \cos\delta) \qquad (1)$$

Unlike actual microphone characteristics, which vary with frequency, the virtual microphones in ViMiC are designed to apply the concept of microphone directivity without simulating undesirable frequency dependencies. The microphone directivity can be continuously changed in real-time. Further, in many spatialization algorithms, sound sources are assumed to be omnidirectional. Because real sound sources are usually directive, ViMiC also features a source directivity model.

To summarize, Inter-Channel Level Differences due to the overall gain between a sound source and the virtual microphones is determined by the source radiation pattern, the distance between source and microphone and the microphone's directivity. Inter-Channel Time Differences are created due to individual distances between a sound source and virtual microphones.
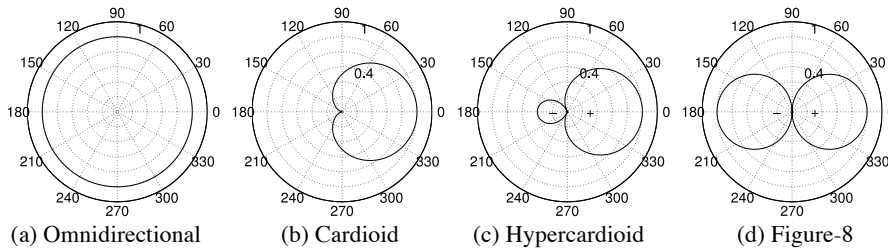

| (a) Omnidirectional | (b) Cardioid | (c) Hypercardioid | (d) Figure-8 |

Figure 3. Examples of common microphone directivity patterns.

2.2 Room Model and Late Reverb

ViMiC contains a shoebox room model to generate time-accurate early reflections that increase the illusion of this virtual space and a sense of envelopment. According to virtual room size and position of the microphones, early room reflections are rendered in 3D through the popular image method of Allen and Berkley (1979). For a general discussion of the image method, see e.g., (Cremer and Müller, 1982). Each image source is rendered according to the time of arrival, the distance attenuation, microphone characteristics, and source directivity, as described in Section 2.1. Virtual room dimensions (height, length, width) alter the reflection pattern accordingly. The spectral influence of definable wall materials can also be simulated.

Early reflections are discretely rendered for each microphone, as propagation paths differ. For five virtual microphones, 35 paths are rendered if the 1st-order reflections are considered (5 microphones · [6 early reflections + 1 direct sound path]). If the 2nd-order early reflections are computed, 95 propagation paths have to be rendered.

Although the reflections are efficiently implemented using shared ring buffers, this can be computationally intensive.

The late reverberant field is very diffuse and ideally without directional information. We synthesize this late reverberant field with a feedback delay network (FDN) structure (Jot and Chaigne, 1991) with 16 modulated delay lines. By feeding the outputs of the previously described room model into the late reverb, an individual, uncorrelated diffuse reverb tail is synthesized for each microphone channel. The tails' timbral and temporal characters can be modified.

Figure 4 depicts a room impulse response recorded within the ViMiC system. Two virtual cardioid microphones were arranged for a stereophonic recording (left side) within a 10m x 15m x 6m virtual room with brick wall reflection pattern. The right panels of Figure 4 show the captured impulse responses for both microphones. Note the Inter-Channel Time- and Level differences of the direct sound component and early reflection pattern in the first 0.1 seconds of the impulse responses. With a different microphone configuration (e.g., a coincident XY microphone setting) the Inter-Channel Time- and Level differences of the room impulse response would be different.
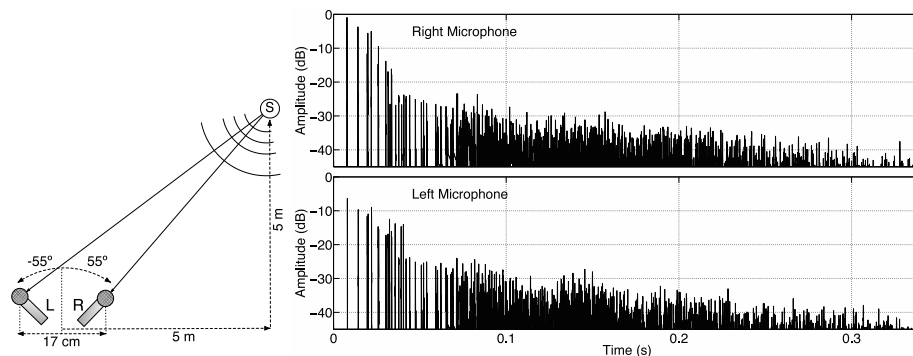

Figure 4. Simulated room impulse response captured with a virtual microphone arrangement of a 10m x 15m x 6m room. The virtual source was positioned five meters to the left, five meters to the front and two meters above the microphone array. The right panel shows the normalized impulse responses of both microphones.

3 Design Approach

Our design approach is strongly based on real-world user needs and behaviours. In an interdisciplinary study, we surveyed spatial sound synthesis techniques across media artists and composers (Peters et al., 2011). Respondents assessed the importance of 10 technical features on a 5-point scale, ranging from "not important" to "extremely

important". This method enabled us to identify the most desired features and then to implement them into the software design process of ViMiC.

The three features with the highest on-average importance ratings were "Spatial rendering in real time", "Controllability via graphical user interface" and "Controllability via external controllers". Integration of the spatial sound renderer through plug-ins into Digital Audio Workstations (DAW) was rated as "extremely important" by a subset of participants. This section discusses the basic ViMiC concept, the findings, and the design and development process that created a flexible spatialization application used for more than media arts and composition.

3.1 Real-time Rendering

To allow fast dynamic control of all parameters in real time, ViMiC deliberately avoids latency-causing blockwise FFT/IFFT spectral filtering processes by employing FIR and IIR filters computed in the time-domain. Therefore, low-latency control of the real-time audio rendering process is supported, which is desired by many composers. For instance, a perceptual event hardly possible in physical environments can now be created by manipulating the virtual room dimensions in real time: the virtual recording room can be transformed from a reverberant cathedral into a small garage-band rehearsal room. Furthermore, according to artistic aspiration or the limitation of the DSP resources, the ViMiC system also offers several levels of rendering quality that can be changed seamlessly.

### 3.2 Environments & Integratability

ViMiC was developed for two popular computer music software paradigms: First, for real-time media programming environments Max/MSP (Zicarelli, 2002); second as a multichannel Audio Unit plug-in for DAWs.

For the Max/MSP development, the Jamoma platform (Place and Lossius, 2006) was used, which provides configurable, easy-to-use higher-level modules with a standardized graphical user interface. The freely available Jamoma distribution includes ViMiC (Figure 5.a) and more than 100 other high-level modules, ranging from controller through video and audio effects to gesture and motion analysis. Hence, ViMiC can be easily combined with and integrated into many different scenarios. Customized stand-alone applications can be created to liberate the application from any software dependency.

The survey indicated that DAWs are very popular environments for spatial audio production. However, DAWs are often tailored to consumer media production and are therefore constrained in their multichannel capabilities (Peters et al., 2009). To make ViMiC accessible in DAWs, a ViMiC multichannel plug-in that can be used with any DAW that supports Apple's Audio Unit plug-in format was developed (Figure 5.b).
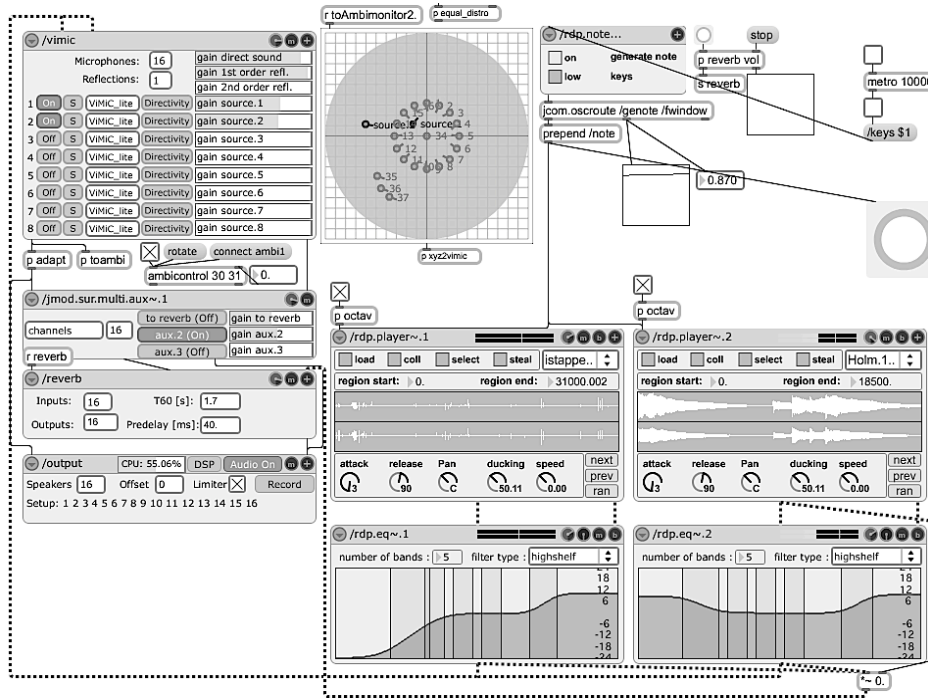
(a)
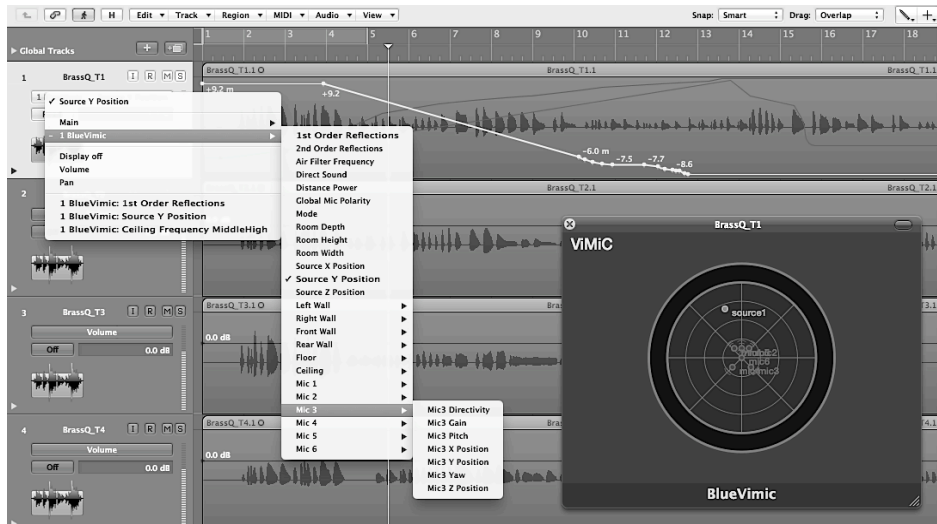
(b)

Figure 5. ViMiC in different audio environments. (a) ViMiC as a Jamoma module (upper left corner) inside Max/MSP. (b) ViMiC as an Audio Unit plug-in inside the audio sequencer Apple Logic.

3.3 Flexible Loudspeaker Settings

Today, as a quasi-standard, many electroacoustic music festivals provide a loudspeaker system based on eight circular, horizontally arranged full-range loudspeakers. For consumer media productions, the standard for surround sound is determined by DVD and Blu-ray Disc, with respectively 5.1 and 7.1 discrete loudspeaker channels[i], according to the ITU (1992) recommendation. Despite these well-known settings, many performance and installation artists have reported using non-standard loudspeaker layouts differing in number, position and elevation of loudspeakers.

ViMiC is currently able to simulate 32 virtual microphone channels. If every microphone is routed to a discrete loudspeaker channel, 32 loudspeakers can be accommodated. Most DAWs usually only support standard surround loudspeaker configurations to a maximum of eight loudspeaker channels sufficient enough for DVD or the emerging Blu-ray Disc.

The ViMiC plug-in is capable of supporting all of these standard configurations. Because the virtual microphones can be freely placed, there is a liberal relation between loudspeaker and virtual microphones, and non-standard loudspeaker settings can be accommodated. Further, the virtual microphone signals might also be post-processed by another spatial rendering or mixing technique e.g., creating a binaural mix for headphone use.

3.4 Accessibility

 The survey showed that new music software applications must be easily integrable into the user's software environments. Accessibility and controllability across different applications are therefore important technical aspects. For instance, the developers of the spatial authoring software MusicSpace (Pachet and Delerue, 1999), winner of the Bourges Music Software prize 2000, mention on their website[ii] that although this stand-alone application received very positive feedback from artists, it was not used more widely because *"MusicSpace was a closed system, not able to communicate easily with other music software"*.

To ease accessibility, besides being equipped with a Graphical User Interface, ViMiC takes advantage of Open Sound Control (OSC, Wright and Freed, 1997), a message protocol for controlling processes across applications and interfaces in real time. The OSC namespace reflects the three main categories (sound source, microphone and room) and is human-readable, avoiding potentially misleading abbreviations (see Listing 1 for an example). It follows herewith the SpatDIF standard for describing spatial sound information to facilitate the exchange of spatial audio scenes (Peters et al., 2012).

Many parameters are defined with unit information, which allows for flexible manipulation of parameters in different measures. For instance, the gain controllers are commonly defined within the MIDI range (0 - 127), but other units can be

declared via OSC messages, e.g., the message `/room/reflection/gain.1 -3.0 db` defines the gain value in decibel.

Listing 1. Open Sound Control namespace example

```
/source.2/position -1.5 4.0 0.0 xyz
/room/reflection/gain.1 100.0
/room/reflection/airfilter 6000 hz
/microphone.5/directivity/preset supercardioid
```

3.5 Perceptual Parameters

Peters et al. (2011) also inquired into the perceptual aspects that composers and artists are striving to create. From fifteen given spatial percepts, the three most highly rated aspects were "Immersiveness", "Distance perception of sound sources" and "Localization accuracy of sound sources".

ViMiC's approach to creating these three perceptual aspects is to simulate the sound propagation in a room as defined through the direct sound, early reflections, and late reverb segments, as previously described. Adapted from Theile (2001), Table 1 shows the contribution of these segments to the perception of envelopment (as related to immersiveness), direction, distance and depth impression. Because early reflections were found to be specifically important for Distance, Depth and Spatial Impression, ViMiC time-accurately renders early reflections for each individual virtual microphone.

Table 1. Contribution of Direct Sound, Early Reflections and Late Reverb to various perceptual aspects. The more stars, the higher the contribution. (adapted from Theile, 2001).

Percept	Direct Sound	Early Reflections	Reverberation
Direction	**	*	
Distance, Spatial Depth		**	*
Spatial Impression		**	**
Envelopment			**
Sound Coloration	**	*	**

4 Case Studies

ViMiC has been applied by a diverse group of users in a variety of acoustical contexts across disciplines. Our case studies include installation artists, composers, sound designers and engineers, many working in film, media, recording and performance arts. ViMiC is also an educational and research tool, in such fields as speech intelligibility and electroacoustic preservation. The studies demonstrate the extent to which our sound spatialization technology meets many needs, and points to future potential uses.

4.1 Musical Applications

Due to its flexibility in the number and placement of virtual microphones and sound sources, ViMiC can be used for non-standardized loudspeaker setups and non-centralized audiences as well as standardized contexts, thus making it applicable for a variety of sound and media applications. The following examples are taken from sound and media installations, concert scenarios, telepresence, studio production, motion picture, and digital preservation of music technology.

4.1.1 Sound and Media Installations

The Wooster Group – There is still time . . . brother In this interactive multimedia piece, the ViMiC system was used in conjunction with a 360◦ cylindrical digital video projection screen (McGinity et al., 2007). The work was commissioned by Rensselaer Polytechnic Institute's Experimental Media and Performing Arts Center (EMPAC, Troy, USA) and presented at the ZKM in Karlsruhe, Germany, and other venues.

The audience is surrounded by the screen with one audience member sitting on a special rotating chair in the central position. Although the projected movie was created for a full 360◦ panoramic screen, only the video segment, which is faced by the person on the rotational chair, is fully visible. The other segments are blurred out to simulate the effect of peripheral field of vision. This "windowed" video segment follows the movement of the rotating chair. Because the timeline of the video is preserved, the audience can explore the entire movie content by viewing it several times from different angles.[iii]

Similarly to the video projection, the accompanying sound field also dynamically adapts in real time to the movement of the rotating chair. ViMiC was used in this scenario to create spatial effects such as depth illusions and Doppler shifts and to simulate various classical microphone techniques. Surreal scenes could be created by assigning artificial directivity patterns to microphones or changing the laws of physics in the model. While the unprocessed audio content and pre-arranged control data were organized on a DAW and streamed to a dedicated ViMiC audio rendering computer, various ViMiC parameters were adjusted in real time, including the directivity patterns and orientations of both the microphones and sound sources, as well as their precise locations. For communication between the rotation chair and the video and audio components, the OSC protocol was used. In total, 54 sound sources were spatialized with ViMiC through 24 loudspeakers. To fully immerse the audience in the installation, these loudspeakers were configured in three rings at different elevations behind the acoustically transparent 360◦ video projection screen, spatializing the sounds, early reflections and late reverb in 3D.

Ricardo del Pozo – adaptation/volume Ricardo del Pozo created the sound installation *adaptation/volume* in partial fulfillment of the requirements for a Master in Fine Arts at the National Academy of the Arts, Bergen, Norway. The 16-channel sound installation was made with the real-time media programming software

Max/MSP. For the installation, pre-recorded sounds were manipulated through different real-time audio effects and spatialized for the 16 loudspeakers via ViMiC. Pozo describes his experience as follows:

> My work deals with the aspect of acousmatics, space and the idea of organized sound, structure and composition as a spatial and sculptural form, how organized sound achieves a body, a form, through spatialization techniques. It is a study into how virtual space overlaps the physical, auditory perception of space and visual perception of space, one superimposed on the other. (Pozo 2010, email communication with the first author).

Compared to the work by the Wooster Group, Pozo used a very different loudspeaker arrangement. Loudspeakers usually surround the audience. He arranged the 16 loudspeakers on a small circle, facing outwards, interacting with the gallery space acoustics (see Figure 6). The ViMiC algorithm was accordingly set up for 16 virtual cardioid microphones on a circle, pointing inwards to the circle's center point. The virtual sound source positions, early reflection pattern, reverberation time and other properties of the virtual room changed very slowly, provoking the audience's acoustic awareness. Pozo said that a key factor in choosing ViMiC for this installation was because the positioning of the speakers was not predefined, ViMiC can be quickly adapted to any loudspeaker setting. *"This means a lot to me since it gives me great flexibility to use this system differently if the space I am working in is architecturally and acoustically challenging."* Prior to using ViMiC, the artist also experimented with Ambisonics and other rendering approaches to create an immersive environment. Below is the comment of Pozo on the subject:

> I found that I never really felt that the perception of depth was audible. When I first tried ViMiC I was struck by how believable it was. It really felt like the sound came from behind the speakers and not coming from the speakers. Also this perception of depth and distance was [..] what I was seeking in my work. (Pozo 2010, email communication with the first author).

Figure 6. Sound installation *adaptation/volume* by Ricardo del Pozo.

4.1.2 Concert Scenarios

Sean Ferguson – Ex Asperis Composed for solo cello, gesture-controlled spatialization, live electronics and chamber orchestra, the world premier of Ex Asperis took place in Pollack Hall at the 2008 MusiMars Festival in Montreal with Chlo´e Dominguez (solo cello), Fernando Rocha (data-gloves for manipulating sound spatialization) and the McGill Contemporary Music Ensemble, directed by Denys Bouliane.

Pollack Hall is a traditional shoebox-style concert venue with raked seatings for up to 600 listeners. The technical setup for this performance was complex. It included bidirectional audio and network connections between stage and front-of-house, on-stage motion-sensors, data-gloves, as well as several decentralized computers used for real-time processing of motion data and spatial-sound rendering. For the spatialization, 24 loudspeakers were arranged in two rings at different heights surrounding the audience. Four additional subwoofers enhanced the low frequency content.

The composer's idea was to use the ViMiC spatialization system to expand the physical limits of the stage width by virtually stretching it during the performance completely around the audience, as controlled by the performers' gestures. To this end, a performer was equipped with data-gloves and binaural headphones to act as a "spatial orchestrator" who arranged and manipulated the spatialized sound sources around the audience, and a sensor system was attached to the right arm of the solo-cellist to measure the activity of bowing motions. The sound of the solo cello was captured live (Figure 7 left) to render virtual early reflections through ViMiC. Played back over the loudspeakers, these reflections enhanced the natural direct sound of the instrument in a subtle, yet perceivable way.

Figure 7. Rehearsal scenes from Ferguson's Ex Asperis. Left: C. Dominguez, equipped with motion sensors in a rehearsal. In front, a feedback-protected microphone for sound capturing for real-time sound spatialization. Courtesy of M. Marshall. Right: View from the Front of House (FoH) to the stage, ViMiC sound processing Max patch on the middle computer screen. Courtesy of R. McKenzie.

The composer prepared a few sound layers for a 7.0 loudspeaker setup using the DAW's built-in panning features. In the concert hall, ViMiC was used to up-mix these 7-channel pre-rendered audio materials to the specific 24 loudspeaker configuration: according to the placement of the 7 loudspeakers in the studio, 7 virtual sound sources were arranged in ViMiC, behaving as virtual loudspeakers. Further, 24 virtual microphones were positioned, according to the placement of the loudspeakers in the hall. By feeding the pre-rendered audio material into ViMiC, the audio was reproduced at the positions of the virtual loudspeakers and "re-recorded" via the virtual microphones.

Marlon Schumacher – De Vive Voix II For Montreal's MusiMars festival 2010, Marlon Schumacher performed his composition *De Vive Voix II* for voice, data glove and live electronics with spatialization of real-time processed sounds in the fairly reverberant Redpath Hall at McGill University. The arrangement of stage and audience seating area, and the limited room size, complicated the standard placement of equidistant surrounding loudspeakers. ViMiC was used in this scenario to develop a creative spatialization concept with eight loudspeakers, arranged as illustrated in Figure 8.
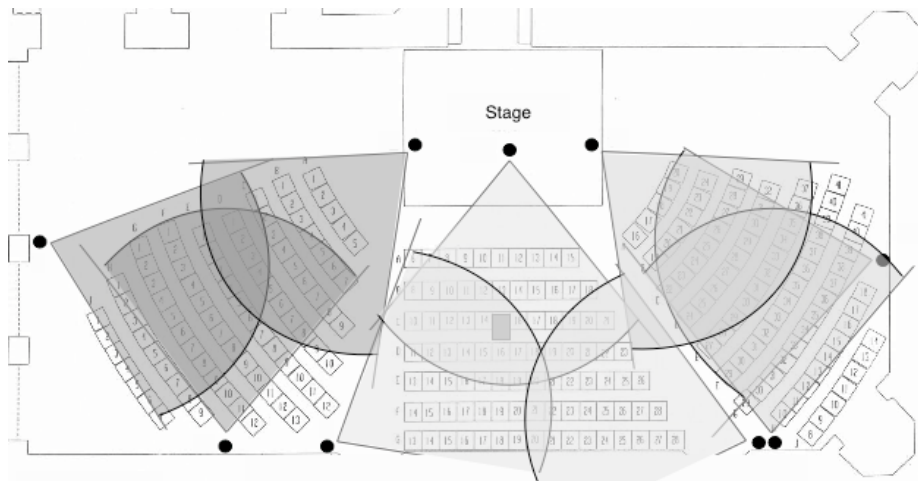

Figure 8. Loudspeaker configuration in Schumacher's *De Vive Voix II*.

The loudspeakers were positioned to provide three separate spatialization and amplification zones, and an overall layer of spatialization, each of which were separately connected to a different ViMiC system. The three zones consisted of the left wing, the right wing, and the central area (Figure 8). Different spatial sound layers were projected with ViMiC to each of the zones, enabling the composer to create "sound clouds", rich in spatial and spectral detail. An overall sound layer used all eight loudspeakers and produced a more "global" sound spatialization, perceivable across the entire audience. Therefore, rather than using the concept of a single ideal listening point (sweet spot) and many low-fidelity listening positions, *De Vive Voix II*

created different listening areas and presented distinct perspectives on the musical material, while providing an overall shared musical experience.

4.1.3 Telepresence Concerts

Telepresence concerts, or live networked music performances, have become popular over the last few years. Increased availability of fast and reliable broadband internet connections and other advances in computer technologies suggests this trend will continue. In these concerts, musicians perform together over the internet, while being physically located at two or more remote sites. In this context, the ViMiC system can be used to create a common auditory virtual space in which all musicians perform and interact with each other (Braasch et al., 2007).

Since 2006, ViMiC has been used for such tasks in telepresence music improvisations between the Tintinnabulate Ensemble (Rensselaer Polytechnic Institute, Troy) and the Soundwire Ensemble (Stanford University) as a component of the Telematic Music System (Braasch, 2009). The Telematic Music System also includes the Expanded Instrument System (Gamper and Oliveros, 1998), JackTrip audio streaming software (Caceres and Chafe, 2009) and the Ultravideo Conferencing system (Cooperstock et al., 2004). In this system (Figure 9), the sound of the musicians is captured using near-field microphones and a microphone array to localize them. The near-field microphone signals are transmitted via JackTrip and spatially recreated at the remote ends using ViMiC and a loudspeaker array. To simulate the same virtual room at all co-located sites, the ViMiC systems communicate using the OSC protocol to exchange room parameters and the room coordinates of the musicians. Using OSC, they also receive localization data from the microphone arrays. An additional bidirectional video stream allows visual interaction among the musicians.
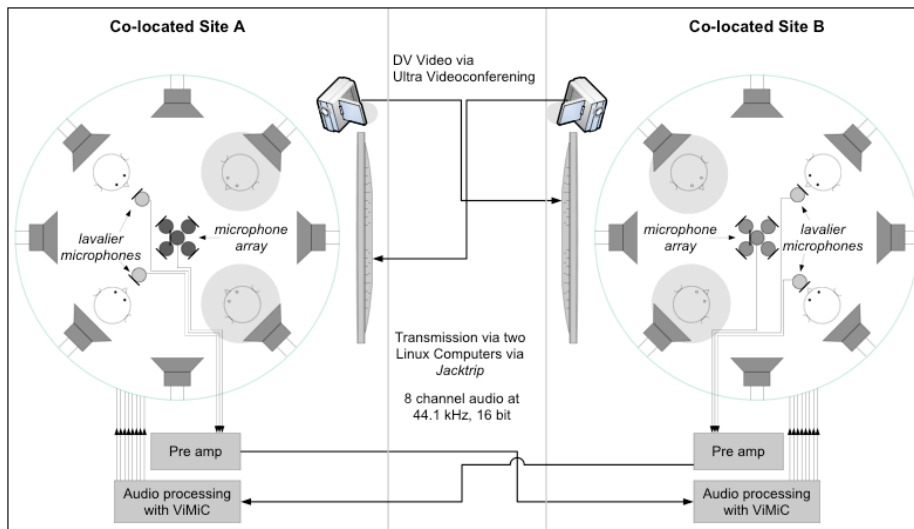

Figure 9. Sketch of the Telematic Music System (Braasch, 2009).

The first commercial album using ViMiC in a telepresence scenario is a 5-channel Quicktime video of a live-recording by the Tintinnabulate and Soundwire ensembles (Tintinnabulate & Soundwire et al., 2009). For the ICAD 2007 conference, they performed *Tele-Colonization* together at the co-located sites McGill University (Montreal, Canada), Rensselaer Polytechnic Institute (Troy, NY, US), Stanford University (Stanford, CA, US), and KAIST (Seoul, South Korea) (Stallmann, 2007).

4.1.4 Studio Production

ViMiC is also used for studio production in commercial audio media formats. For example a two-channel auralization can be heard on the CD Global Reflections by Jonas Braasch (2006) and on a 5.1 DVD of Marlon Schumacher's composition De Vive Voix II. To arrange the virtual microphones for commercial media formats (e.g., stereo, ITU 5.1), Tonmeisters developed different microphone setups (for an overview see Williams and Le Duˆ, 2004; Rumsey, 2001) that are easily applicable in ViMiC.

Because audio productions usually employ DAWs, the ViMiC Audio Unit plug-in can be used. In a DAW, the ViMiC plug-in is added to the desired (to be spatialized) audio tracks and, by manipulating the plug-in's GUI, the positions of microphones and sound sources are defined. All parameters can be dynamically controlled via the DAW's automatization features in real time. To facilitate the microphone setup, many common microphone techniques are available as factory presets. This flexibility uniquely positions ViMiC in the context of radio play productions, where dynamic changes of listening perspective and room environment are prominent dramaturgical principles.

In the audio production context, ViMiC provides great flexibility. Imagine a sound engineer who has completed a multichannel (multi-microphone) recording of an orchestra. Then the audio producer decides to add an extra sound layer on top of this recording. Usually, a simple amplitude panning would be used to position and distribute the sounds of the extra layer on top of the recording. Because of the missing Inter-Channel Time Differences in the added sound layer, the mix with the previously recorded orchestra may sound flawed. By arranging ViMiC's virtual microphones and room parameter similarly to that of the real recording, the extra sound layer can be spatially matched to the recorded material, thus creating a more homogenous spatial sound impression than could have been achieved by simple amplitude panning. Also in the context of mixed music productions (music for acoustic instruments and electronics), ViMiC can help to blend electronic sounds with those of the acoustical instruments, an often-desired effect.

4.1.5 Digital Preservation of Stockhausen's Rotation Table

The preservation of electroacoustic music is becoming an important topic among composers, musicians, musicologists and researchers in Information Studies and Music Technology (Chadabe, 2001). Several efforts have been made to digitally recreate old technology for analysis and experiential purposes (e.g., Clarke and

Manning, 2009). The fast development and generation changes in media formats and music technology complicate the issue, making technology often obsolete and inaccessible even before its musical potential can be fully explored.

One of these technologies is the Rotation Table, which was developed in 1958 by Karlheinz Stockhausen for his piece Kontakte and later refined for Sirius (1975–77). A directional and rotatable loudspeaker was surrounded by four stationary microphones that receive the loudspeaker signal (Figure 10.a). The recorded microphone signals were played back and routed to different loudspeakers arranged around the audience. Due to the directivity and separation of the microphones, the recorded audio signals contained Inter-Channel Time Differences (ICTDs) and Inter-Channel Level Differences (ICLDs). The speaker could be manually rotated up to about 7 Hz and, depending on its velocity, the change in ICTDs created audible phasing and Doppler effects. For high rotation frequencies, the sound "starts dancing completely irregularly in the room – at the left, in front, it's everywhere"– even changing pitch depending on where the listener is standing. This is caused by the phase-shifting effects, which alternately stretch and compress the sound. Stockhausen believed that such effects could not be reproduced by simple amplitude panning (Maconie, 2005).

ViMiC was used to emulate the loudspeaker-microphone configuration of the rotation table to produce the ICLDs and ICTDs that create the sound quality as reported by Stockhausen. Figure 10.b shows the application where pre-recorded or live sounds can be spatialized in real time. The table's rotation speed, direction, and the loudspeaker distance from the table's rotation axis can be manipulated either via the GUI, or remotely with an external controller. Consequently, the rotation table technology is digitally preserved and is reusable with a variable number of virtual microphones for today's musical applications.
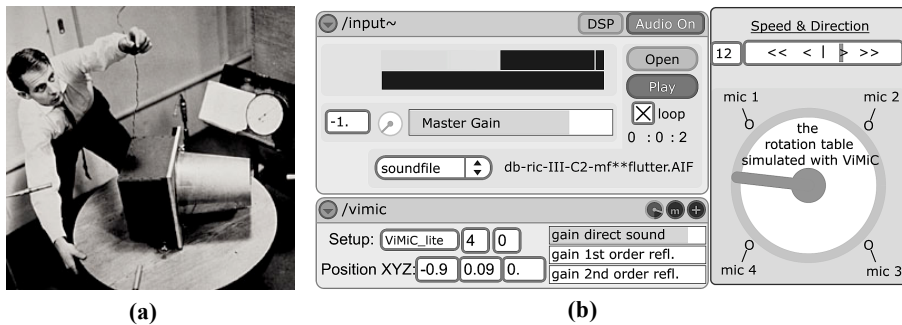

(a)   **(b)**

Figure 10. Preservation of Stockhausen's Rotation Table using ViMiC: (a) Stockhausen with his table, (b) The digital emulation using ViMiC. © left: Stockhausen Foundation for Music, Kürten, Germany (www.stockhausen.org)

4.1.6 Motion Picture

Spatial sound is a popular creative element in Cinema and Motion Picture, where the interaction between picture and audio can create a unique experience for the audience. Psychological studies have shown that sound in combination with visual stimulation is better at generating entertainment pleasure and emotions than visual inputs alone (Christensen and Lund, 1999).

Molecules to the MAX! Premiered in 2009, the 3D-animation movie for IMAX *Molecules to the MAX!* is an educational family adventure where molecules, the main characters, travel with a spaceship through the universe. On their journey, the cartoon figures explore on a microscopic level, the molecular world of snowflakes, raindrops, and other objects.

Sound designer Jesse Stiles used an early version of the ViMiC Audio Unit plug-in to spatialize sounds. From the visual rendering software Maya, used to create the animations, he received a real-time OSC data stream containing the spatial locations of the animated characters and sounding objects in Maya's virtual world. These position coordinates were then used to spatialize sound effects and dialog with the ViMiC plug-in for the six-channel IMAX surround format synchronous to the picture. Further, by properly placing the virtual microphones in ViMiC, the typical mismatch between the screen width and the width spanned by the frontal loudspeakers was compensated for.

The virtual microphones were also oriented according to the camera perspective. Whenever the camera perspective changed, the virtual microphones automatically displaced and oriented themselves according to this real-time synchronization with Maya. For instance, a 360º panning shot makes the virtual microphones rotate simultaneously with the camera. With virtual camera position and virtual microphone positions always in synchrony, the time-consuming need to create sound trajectories manually according to the camera perspective was eliminated.

The multichannel audio tracks were sent to the Technicolor studio in Toronto for the final sound mix. Technicolor's sound engineers reported that ViMiC's spatialization approach is a novel concept in the context of large-format films and *"seems to work well with the image"*.

4.2 Education

4.2.1 Tonmeister Training

Tonmeister students usually undergo technical ear training courses to sharpen perception and understanding of sound quality to improve their recording and production skills. Timbral aspects and reproduction artifacts (e.g., bit-errors, amplitude/phase response differences between channels) are often prioritized in training, with less emphasis on spatial sound attributes (Neher, 2004). ViMiC has the potential to be the missing educational tool for training recording engineers and

Tonmeisters. With ViMiC, students can create virtual recording scenarios and quickly experience the subtle differences between microphone directivities, various microphone techniques, the perceptual differences between Inter-Channel Time Differences and Inter-Channel Level Differences, and the effect of source directivity patterns and room reflections. Because ViMiC is designed as a real-time application, having a constrained amount of processing power, the software does not simulate frequency-dependent directivity characteristics of specific microphone brands and models. However, with faster computer systems, this feature may be added to make ViMiC even more suitable for this application. ViMiC is equipped with a preset database to simulate popular stereophonic and multi-channel microphone settings including XY, Decca, or Fukada Tree (see Figure 11), to cover a variety of multichannel microphone techniques. The preset database addresses the consensus among sound engineers that there is no paramount or ideal microphone configuration for all possible recording and listening scenarios.
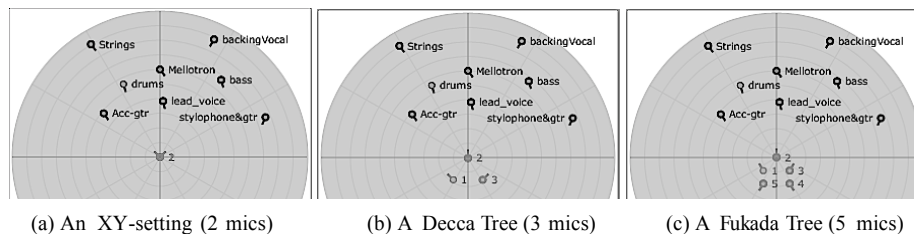

(a) An XY-setting (2 mics)   (b) A Decca Tree (3 mics)   (c) A Fukada Tree (5 mics)

Figure 11. Top view of a ViMiC recording scene, using different microphone arrangements. Dots with text label: sound sources and their frontal direction. Numbered dots: microphones. The "nose" on the circles illustrates the orientation of the sound sources and microphones.

4.2.2 Educational Events

ViMiC was featured by the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) at several educational events primarily for children. An interactive spatial sound installation using between 8 and 22 loudspeakers with a multitouch interface (Figure 12) enabled visitors to manipulate and explore the spatial location of multiple sound sources in different sound scenes. For instance, one sound scene involving an orchestra divided into eight instrumental sections performing a Beethoven symphony allowed participants to experience a virtual concert hall scenario. Children used the interface to listen to the acoustic character of a particular section, and to virtually position orchestra sections in different spatial formations. To account for different age groups, composer and sound designer Eliot Britton created a variety of sound scenes, including an urban environment, a rain forest, and a rock concert. Figure 12 shows a child at the *Eureka!* science festival discovering and virtually arranging a soundscape with a multitouch interface connected via OSC to the ViMiC system.

Figure 12. At the Eureka! Science Festival. Left: Loudspeakers and the computer system running the ViMiC system. Right: Multitouch interface.

4.3 Research Projects Using ViMiC

4.3.1 Medical Sector

ViMiC is currently used at the Boys Town National Research Hospital in Omaha, NE, USA to assess children's speech intelligibility in the presence of noise and reverberation by using a virtual classroom paradigm (Valente et al., 2012). For the experiments, audio/visual stimuli of children reading classroom lessons were created and processed with the ViMiC system to generate controlled room models. By changing a number of ViMiC parameters, room acoustical properties of these models were varied. Background noise to simulate ventilation and air conditioning was added with varying levels.

4.3.2 Sound Recording Research

With the advent of new consumer surround reproduction standards ranging up to 22.2 (Hamasaki et al., 2004) loudspeaker systems, traditional five-channel recording techniques used for the 5.1 standard are insufficient. Consequently, adequate multichannel microphone techniques have to be developed and evaluated. ViMiC can help in this process: new microphone configurations can be virtualized and tested before time- and money-consuming recordings in real-world scenarios are created.
Braasch et al. (2009) included ViMiC in a specifically designed mixing console for telematic music. This new mixing application includes a number of specific features for telematic music such as a meter to measure the latency and sound level between the remote venues. ViMiC is used here to spatialize the incoming spot-microphone recordings from the remote venues.

Parks and Braasch (2011) used ViMiC in listening tests to investigate the roll of head movements in the perception of spatial sound. The tests focused on the two aspects of

spaciousness "Listener Envelopment" (LEV) and "Apparent Source Width" (ASW). Results show that head movements are critical for the perception of the ASW. No effect was found for LEV.

5 Conclusion

This paper has shown how an interdisciplinary approach helped to effectively refine the design and development of the ViMiC spatialization system. This approach included a survey to understand user's needs and priorities, studies of user scenarios, and real-world test cases. To make ViMiC accessible to other users, we are planning to make ViMiC available for other computer music software environments such as Super Collider or Pro Tools.

We hope that ViMiC can contribute to the exploration of spatial sound characteristics and that our development approach will guide other efforts to create relevant, user-friendly tools.

6 Acknowledgment

References

Allen, J. B. and D. A. Berkley (1979). Image method for efficiently simulating smallroom acoustics. *J. Acoust. Soc. Am. 65 (4)*, 943–950.

Berkhout, A. J., D. de Vries, and P. Vogel (1993). Acoustic control by wave field synthesis. *J. Acoust. Soc. Am. 93*, 2764–2778.

Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. Cambridge, Mass.: MIT Press.

Board, P. (2002). Biography Henry Brant. http://www.pulitzer.org/biography/2002-Music, accessed August 2012.

Borwick, J. (1973, 9). The Tonmeister concept. In *Proc. of the 46th AES Convention, Preprint 938*.

Braasch, J. (2005). A binaural model to predict position and extension of spatial images created with standard sound recording techniques. In *Proc. of the 119th AES Convention, Preprint 6610*, New York, NY, USA.

Braasch, J. (2006). *Global Reflections*. Kingston, US: Deep Listening DL 34-2006. Braasch, J. (2009). The telematic music system: Affordances for a new instrument to shape the music of tomorrow. *Contemporary Music Review 28 (4)*, 421–432. Braasch, J., C. Chafe, P. Oliveros, and D. V. Nort (2009). Mixing-console design considerations for telematic music applications. In *Proc. of the 126th AES Convention, Preprint 7942*, New York, US.

Braasch, J., N. Peters, and D. L. Valente (2007). Sharing acoustic spaces over telepresence using virtual microphone control. In *Proc. of the 123rd AES Convention, Preprint 7209*, New York, US.

Braasch, J., N. Peters, and D. L. Valente (2008). A loudspeaker-based projection technique for spatial music applications using virtual microphone control. *Computer Music Journal 32 (3)*, 55–71.

Caceres, J.-P. and C. Chafe (2009). Jacktrip: Under the hood of an engine for network audio. In *Proc. of the International Computer Music Conference*, Montreal, Canada, pp. 509–512.

Chadabe, J. (2001). Preserving performances of electronic music. *Journal of New Music Research 20 (4)*, 303 – 305.

Christensen, K. B. and T. Lund (1999). Room simulation for multichannel film and music. In *Proc. of the 107th AES Convention, Preprint 4993*, New York, US.

Clarke, M. and P. Manning (2009). Valuing our heritage: exploring spatialization through software emulation of Stockhausen's Oktophonie. In *Proc. of the International Computer Music Conference*, Montreal, Canada, pp. 179–182.

Cooperstock, J. R., J. Roston, and W. Woszczyk (2004). Broadband networked audio: Entering the era of multisensory data distribution. In *Proc. of the 18th International Congress on Acoustics*, Paris, France.

Cremer, L. and H. A. Müller (1982). Principles and Applications of Room Acoustics, (translated by T. J. Schultz). *Applied Science Publishers 1*, 17–19.

Galeyev, B. (2006). Spatial music. http://prometheus.kai.ru/pr-mys_e.htm, accessed August 2012.

Gamper, D. and P. Oliveros (1998). A performer-controlled live sound-processing system: New developments and implementations of the expanded instrument system. *Leonardo Music Journal 8 (1)*, 33–38.

Gerzon, M. A. (1973). With-height sound reproduction. *J. Audio Eng. Soc. 21 (1)*, 2–10.

Hamasaki, K., S. Komiyama, H. Okubo, K. Hiyama, and W. Hatano (2004). 5.1 and 22.2 multichannel sound productions using   an integrated surround sound panning system. In *Proc. of the 117th AES Convention, Preprint 6226*, San Francisco, US.

ITU (1992). *Recommendation BS.775-2, Multichannel stereophonic sound system with and without accompanying picture*, Geneva, Switzerland: International Telecommunication Union.

Jot, J. and A. Chaigne (1991). Digital delay networks for designing artificial reverberators. In Proc. of the 90th AES Convention, Preprint 3030, Paris, France.

Klapholz, J. (1991). Fantasia: Innovations in sound. *J. Audio Eng. Soc. 39* (1/2), 66–70.

Maconie, R. (2005). *Other Planets: The Music of Karlheinz Stockhausen*. Scarecrow Press.

Malham, D. G. and A. Myatt (1995). 3-D sound spatialization using Ambisonic techniques. *Computer Music Journal 19* (4), 58–70.

Marshall, M., N. Peters, A. Jensenius, J. Boissinot, M. Wanderley, and J. Braasch (2006). On the Development of a System for Gesture Control of Spatialization. In *Proc. of the International Computer Music Conference*, New Orleans, US, pp. 360–366.

McGinity, M., J. Shaw, V. Kuchelmeister, A. Hardjono, D. D. Favero, and A. Hardjono (2007). AVIE: a versatile multi-user stereo 360◦ interactive VR theatre. In *Proc. of the Workshop on Emerging Displays Technologies*, New York, US.

Moore, B. C. J. (2012). *An Introduction to the Psychology of Hearing* (6th ed.). Emerald Group Publishing.

Neher, T. (2004). *Towards A Spatial Ear Trainer*. Ph. D. thesis, School of Arts, University of Surrey, UK.

Pachet, F. and O. Delerue (1999). Musicspace: a constraint-based control system for music spatialization. In *Proc. of the International Computer Music Conference*, Beijing, China, pp. 272–275.

Parks, A. and J. Braasch (2011). The effect of head movement on perceived listener envelopment and apparent source width. In *Proc. of the 131st AES Convention, Preprint 8567*, New York, US.

Peters, N., T. Lossius, J. Schacher, P. Baltazar, C. Bascou, and T. Place (2009). A stratified approach for sound spatialization. In *Proc. of the 6th Sound and Music Computing Conference*, Porto, PT, pp. 219–224.

Peters, N., T. Lossius, and J. C. Schacher (2012). *SpatDIF: Principles, specification, and examples*. In 9th Sound and Music Computing Conference (SMC), Copenhagen, DK.

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

Peters, N., T. Matthews, J. Braasch, and S. McAdams (2008). Spatial Sound Rendering in Max/MSP with ViMiC. *In Proc. of the International Computer Music Conference*, Belfast, UK, pp. 755–758.

Place, T. and T. Lossius (2006). Jamoma: A modular standard for structuring patches in Max. In *Proc. of the International Computer Music Conference*, New Orleans, US, pp. 143–146.

Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc. 45* (6), 456 – 466.

Rumsey, F. (2001). *Spatial Audio*. Oxford, UK: Focal Press.

Stallmann, K. (2007). Songs Inside My Head: ICAD 2007. *Newsletter of the Society for Electro-Acoustic Music in the United States 4*, 14–15.

Theile, G. (2001). Multichannel Natural Music Recording Based on Psychoacoustic Principles. In *Proc. of the AES 19th International Conference on Surround Sound: Techniques, Technology, and Perception*, Schloss Elmau, Germany, pp. 201–229.

Tintinnabulate & Soundwire, J. Braasch, C. Chafe, P. Oliveros, and B.Woodstrup (2009). *Tele-Colonization*. Kingston, US: Deep Listening DL-TMS/DD-1.

Valente, D., H. Plevinsky, J. Franco, E. Heinrichs-Graham, and D. Lewis (2012). Experimental investigation of the effects of the acoustical conditions in a simulated classroom on speech recognition and learning in children. *J. Acoust. Soc. Am. 131*, 232–246.

Williams, M. and G. Le Duˆ (2004). The Quick Reference Guide to Multichannel Microphone Arrays, Part 2: using Supercardioid and Hypercardioid Microphones. In *Proc. of the 116th AES Convention, Preprint 6059*, Berlin, Germany.

Wright, M. and A. Freed (1997). Open Sound Control: A New Protocol for Communicating with Sound Synthesizers. In *Proc. of the International Computer Music Conference*, Thessaloniki, Greece, pp. 101–104.

Zicarelli, D. (2002). How I learned to love a program that does nothing. *Computer Music Journal 26* (4), 44–51.

Zvonar, R. (1999). A history of spatial music. *eContact!* (7.4).

---

[i] The number behind the dot symbolizes the number of discrete subwoofer channels.

[ii] http://www.csl.sony.fr/~pachet/musicspace.html, accessed Aug 2012

[iii] http://www.icinema.unsw.edu.au/projects/there-is-still-time-brother, accessed Aug 2012.

Biographies

Nils Peters is a postdoctoral fellow at the International Computer Science Institute (ICSI) and the Center for New Music and Audio Technologies (CNMAT) at UC Berkeley. He holds a MSc degree in Electrical and Audio Engineering from the University of Technology in Graz, Austria and a PhD in Music Technology from McGill University in Montreal, Canada. He has worked as an audio engineer in the fields of recording, postproduction and live electronics and is currently working on real-time algorithms for sound-field analysis with large-scale microphone arrays.

Jonas Braasch is a musicologist and aural architect with interests in technologized improvised music, telematic music and intelligent music systems. He studied at the Universities of Bochum and Dortmund (Germany) and received Ph.D. degrees in Musicology and Engineering. He currently works as Associate Professor in the School of Architecture at Rensselaer Polytechnic Institute, where he directs the Communication Acoustics and Aural Architecture Research Laboratory (CA3RL). His work on Telematic Music and Sound Spatialization Systems has received funding from the U.S. National Science Foundation and the Natural Sciences and Engineering Research Council of Canada. He has organized and participated in numerous international telematic music performances.

Stephen McAdams studied music composition and theory before entering the realm of perceptual psychology. In 1986, he founded the Music Perception and Cognition team at the world-renowned music research centre Ircam in Paris. While there he organized the first Music and the Cognitive Sciences conference in 1988, which subsequently gave rise to the three international societies dedicated to music perception and cognition, as well as the International Conference on Music Perception and Cognition. He was Research Scientist and then Senior Research Scientist in the French Centre National de la Recherche Scientifique (CNRS) from 1989 to 2004. He has taken up residence at McGill University since 2004 where he is Professor and Canada Research Chair in Music Perception and Cognition. He directed the Centre for Interdisciplinary Research in Music, Media and Technology (CIRMMT) from 2004 to 2009.