Adaptive Real-Time Spatialization for Live Music Improvisation

Michael Ott*
Conservatorium van Amsterdam
Amsterdam University of the Arts
Amsterdam, The Netherlands
mail@michaelmariaott.com

Atser Damsma
Conservatorium van Amsterdam
Amsterdam University of the Arts
Amsterdam, The Netherlands
atser.damsma@ahk.nl

Abstract

Live spatialization poses a challenge for instrumentalists, because it requires real-time control over numerous parameters. We present a novel system for automatic audio diffusion that is tailored to improvising musicians. The system enables real-time spatialization of acoustic instruments and other monophonic sound sources by dynamically decomposing incoming audio into spectral streams using Non-Negative Matrix Factorization (NMF). By mapping these spectral streams to separate channels, the system transforms single instruments into spatialized soundscapes with emergent movement. Unlike existing methods that rely on offline training or fixed frequency bands, this approach adapts continuously to the improvised material without offline training, while remaining controllable through a single spread parameter. We implemented a Max for Live device that enables performers to easily integrate the system into their improvisational practice, thereby bridging the gap between live acoustic performance and electroacoustic spatial practices.

1 INTRODUCTION

As loudspeaker systems with a larger number of loudspeakers are becoming more common in recording studios and music venues, more music pieces are being written and produced for these systems, providing listeners with an experience that would not be possible when listening to a stereo recording of the material at home (Austin and Smalley, 2000). These pieces generally are either composed or produced specifically for a spatial audio system, in which case the spatialization is part of the composition, or made up of composed or pre-produced material that is spatialized during the performance.

Spatialization of sound during a live performance is done either by controlling the level of the sound sent to each available speaker or by assigning the sound a virtual position in the room and calculating the corresponding levels and filtering for each speaker based on that position (Peters et al., 2011; Frank et al., 2015; Pulkki, 1997; Lossius and Baltazar, 2009). Of course, these systems for spatialization can also be used on improvised material, but they require real-time control of many parameters, whether it is the individual levels of the signal sent to each speaker or the positions of multiple sound sources in the performance space. Even with dedicated controllers (Cannon and Favilla, 2010), this would require one or more performers to focus solely on controlling the spatialization, or the musicians would have to alternate between playing their instruments and controlling the spatialization.

Automating the parameters prior to performance is not an option for improvised performances, because the content and duration of the music is not known beforehand. To expand their expressive possibilities through spatialization, improvising artists would therefore need a system that can automatically diffuse the signal based on the content of the audio input, while still offering enough control and flexibility to be adapted to the artist's setup and improvisational intent.

Spatializing audio with little or no active control is not a new idea. Circumspectral diffusion divides the incoming audio into fixed frequency bands and assigns each frequency band its own location in the physical space (Khosravi, 2011). The frequency-location assignment stays fixed throughout the piece, movement is generated only by the changing energy of each frequency region in the audio input. Since distributing the frequency bins across different loudspeakers introduces artifacts and blurred transients, the bins can only be distributed sequentially to minimize the distance between adjacent bins (Khosravi, 2011). Any movement in circumspectral diffusion is therefore directly related to movement in the spectral domain and does not necessarily correspond to musically separate objects.

Existing solutions for real-time audio decomposition typically require offline training for a predefined number and type of components, are often limited to a small number of components, and are trained to separate audio into individual instruments or voices (Watcharasupat and Lerch, 2024; Sarkar, 2024; Yun-Ning et al., 2025; Petermann et al., 2022; Watcharasupat et al., 2024), which limits their usefulness for improvised live performances.

This paper presents a system for the adaptive spatialization of live improvisation. Using Non-Negative Matrix Factorization (NMF), the input signal is decomposed into a user-defined number of components in real time, without prior offline training. By separating the audio into spectral streams, the spatialization dynamically responds to the evolving sound throughout the performance, while maintaining intuitive control by an instrumentalist. The system was developed for solo saxophone improvisations, but should be flexible enough to be used with other instruments, especially ones that are able to produce both sustained and percussive sounds and a wide range of frequencies.

*www.michaelmariaott.com

2 CONSIDERATIONS FOR AUTOMATED REAL-TIME SPATIALIZATION

For an automated solution to real-time spatialization to be successful, artistic goals as well as technical necessities need to be considered.

2.1 Sound Spaces

To be able to devise a system that expands the expressive possibilities during improvisation, we need to be aware of the spaces that make up the physical environment of a performance.

The physical spaces relevant for a performance range from the gestural and ensemble space (Figure 1a), defining the area immediately around the performer or performers, to the arena space (Figure 1b), enveloping both the performer(s) and the audience (Smalley, 2007).

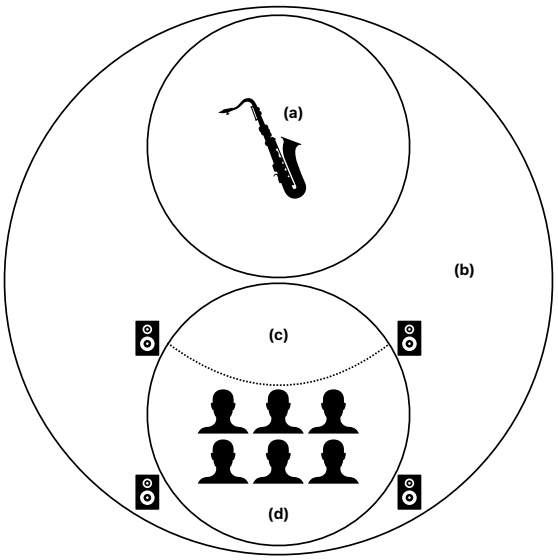

Figure 1: Physical areas in a solo performance situation are (a) the gestural space, (b) the arena space, (c) the prospective space and (d) the circumspace.

If a sound is strongly associated with a particular source, listeners will localize it to that source, even if it originates from different locations (Smalley, 2007, 1997). On the one hand, this limits the possibilities of effectively spatializing live instruments, since the listening of the audience is always going to be centered on the performer (Austin and Smalley, 2000). On the other hand, the expectations created by the presence of a specific instrument also open up opportunities for creating new spatial perceptions by spatializing sounds that fall outside these expectations, such as extended techniques, amplified key noises, or electronically processed instrument sounds.

Perspectival space denotes the acoustic and visual representation of the performance environment as experienced from the listener's position (Smalley, 2007). It is made up of the prospective space (Figure 1c), which unfolds like a landscape with changing width and depth in front of the listener and the circumspace (Figure 1d), which encompasses the listener (Smalley, 2007).

Especially interesting for the spatialization of a single instrument is the distribution of the audio stream across the circumspace. If the decomposition and distribution of the sound source can be controlled in real-time either through minimal direct controls or through audio analysis, this would allow the improviser to use the circumspace to expand the range of expressive possibilities during their improvisation.

2.2 Spatialization as Enhancement

If the music is composed for a specific spatial audio setup, the composer can compose the spatialization together with the music. If the music is composed and later spatialized live, the performer has the opportunity to analyze the spaces and motions implied by the composition and base his spatialization on the analysis. For improvised music however, the improviser is generating material and therefore the spaces derived from the material in real-time. The spectral content of the improvised material has to be interpreted and spatialized as the material is generated.

While it is unrealistic for a performer to keep track of multiple sound sources and their spatial positions and trajectories in real-time (Austin and Smalley, 2000), the improviser may have an intuition about the spatiality that can be derived from the material they are improvising as it is happening. Therefore, it makes sense to base the spatialization on the spectral properties of the improvised material, but also provide the performer with a certain amount of control to communicate that intuition to the algorithm. In the proposed system, the order of the spectral objects can be controlled based on their spectral properties, while a single expression pedal provides control over the strength of the spatialization. In this way, the sound can be diffused across a larger space or confined to its source.

2.3 Movement

Even if the spatialization of the sound remains static and the same frequency ranges or spectral streams always occupy the same position in the space, the continuous movement of the sound in the frequency domain results in perceived motion across the space for the listener (Khosravi, 2011). Therefore, an algorithm to decompose and diffuse an instrument can focus on the creation of the environment from the improvised material and let the change of the material in the frequency domain determine its movement as an emergent feature. The improviser thus has explicit control over the environment by changing the amount of spatialization, and implicit control over the movement by varying the frequency content of the material.

If the source audio is continuously re-analyzed and decomposed, the detected spectral streams will change due to variations in the improvised material, and so will their position in the space, introducing another element of movement.

2.4 Flexibility

There is no industry standard for speaker setups or multi-speaker panning algorithms for performance venues. Varying electroacoustic setups directly affect which properties of the sound are exaggerated or masked (Austin and Smalley, 2000).

Electroacoustic music, as a compositional genre, is often composed independently of the performance space and has to be adapted for the particular setup of each performance

(Austin and Smalley, 2000). Even if a piece was composed for one specific performance space, it will have to be adapted eventually for performances in different locations.

Improvised material is generated on the spot in the performance space and therefore the artist will ideally take the electroacoustic conditions in the performance space into account. To extend that flexibility to the spatialization of improvised material, the algorithm has to be able to adapt to the performance space as well: it should be able to be employed with any speaker setup. Instead of creating precise movement directories, which cannot be reproduced on speaker setups with a lower number of speakers (Catena and Frisk, 2024), it is more important that the sounds move among the available speakers at all.

3 SYSTEM FOR REAL-TIME SPATIALIZATION

3.1 Real-Time Decomposition

We propose a system that can automate real-time spatialization from mono sound sources using audio decomposition, without the need of extensive offline training. Spectral components are detected with Non-negative Matrix Factorization (NMF), an unsupervised learning algorithm for matrix decomposition that has become popular for audio source separation (Lee and Seung, 2000; Févotte et al., 2018). The system analyzes the incoming audio signal (*training*) and filters the spectral components from the audio signal in real-time (*inference*). The re-synthesis of the spectral components can be used to determine their placement in the space based on their spectral properties.

3.1.1 Non-Negative Matrix Factorization

Given a non-negative matrix V, NMF is used to find two matrices W and H so that $V \approx WH$ (Figure 2)(Lee and Seung, 2000). W and H are found by recursively multiplying them by a factor that depends on the quality of the approximation, expressed by a cost function. Repeated iteration guarantees convergence to a local minimum (Lee and Seung, 2000).

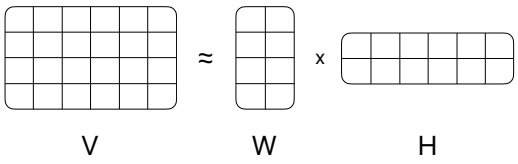

Figure 2: Non-negative matrix factorization.

Applied to audio, V is the time-frequency representation of the audio and after convergence, W contains the frequency content for the spectral streams of the input audio (*bases*) and H their amplitude over time (*activations*) (Févotte et al., 2009, 2018).

For the proposed system, we use the FluCoMa package (Tremblay et al., 2022) for Max (Cycling74, 2026), which implements multiple NMF-based objects for audio processing with a multiplicative update function based on the Kullback-Leibler divergence (Roma, Gerard et al., 2020).

3.1.2 Real-Time Training

The incoming audio is recorded into two alternating buffers of user-defined duration (Figure 3a). While the first buffer is being written to, the second is used for training and vice versa. This makes sure that there are no read/write conflicts and the training material is always one continuous section of audio.

At the beginning of each training, the buffers for the bases and activations are seeded using Non-negative Double Singular Value Decomposition (NNDSVD) (Figure 3b) (The FluCoMa Project, 2022a). This leads to the NMF converging faster, producing better results in the given amount of iterations (Boutsidis and Gallopoulos, 2008).

After the initialization, the user-specified number of bases are generated on a background thread without disrupting the live audio (Figure 3c). The detected spectral streams from the ring-buffer are analyzed offline for their spectral properties (Figure 3d).

3.1.3 Offline Training

In the training method described above, the bases are inferred in real-time during performance. However, if the performer knows the material they are going to use during their performance, they can choose to train the NMF on the material beforehand and only run the inference in real time.

3.1.4 Real-Time Inference

During inference, NMF is performed on a single spectral frame of the audio, using the provided bases as seeds (Figure 3e) (The FluCoMa Project, 2022b). The real-time audio input is filtered into spectral streams based on the bases buffer and bundled into a multichannel signal, based on the amount of spectral objects. After each analysis, there is a crossfade from the old to the new filtered audio signals, the length of which can be set by the user.

If the base for a spectral object is similar to a base in the previous buffer and assigned to a different channel, the crossfade is perceived by the audience as movement between these two channels. The length of the buffer and the crossfade time can be adjusted to either achieve subtle continuous movement or choppy, rhythmic jumps.

3.2 Control

In a basic setup, the order of the generated bases, and therefore the resulting movement in the soundscape, cannot be predicted. In the current system, however, the bases can be sorted based on properties of the objects, such as their spectral centroid (Figure 4). This is helpful if channels should be processed differently based on a certain property. Since the content of the bases cannot be predicted, the values of the given property of each base are compared and the bases are sorted relative to each other. A common application of this is to send channels with higher spectral centroids to the ceiling speakers in a speaker setup with multiple vertical layers, but it could also be used to determine the effects processing of each channel based on its spectral properties (e.g., adding reverb to high-frequency material).

The performer can also control the overall spread of the spatialization in real-time with a controller or expression pedal (Figure 3f). With the spread at zero, all channels are summed to a mono signal and sent to the center speaker(s). With increasing spread, the channels are panned further apart until each channel is sent to a separate speaker.

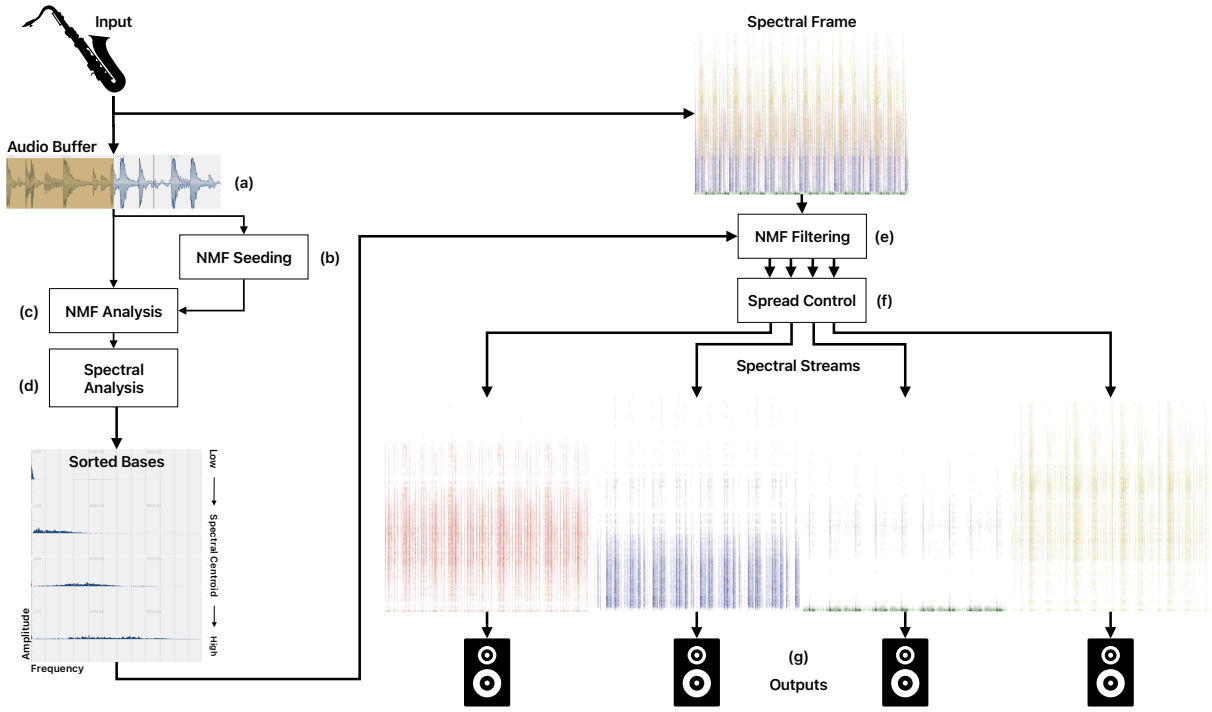

Figure 3: Workflow of the Non-Negative Matrix Factorization audio processing pipeline: Recording the input (a), analyzing (b - d) and filtering (e) it using NMF, before controlling the spread (f) and sending the spectral streams to separate outputs (g).

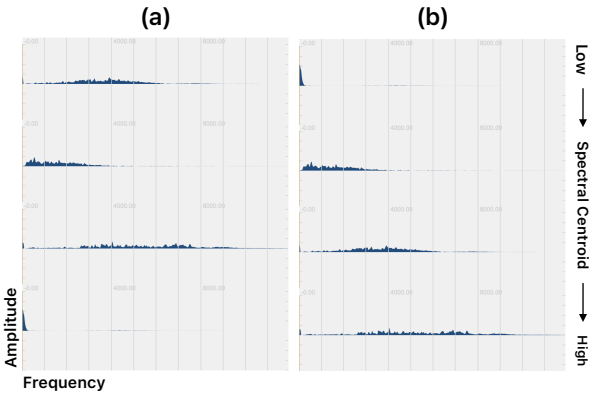

Figure 4: Contents of the bases buffer before (a) and after (b) sorting by mean spectral centroid.

3.3 Max for Live Device

We implemented a Max for Live device based on the described system (Figure 5). The device exposes all settings of the algorithm to the user: the number of desired components, analysis length, crossfade length between decompositions and the spread factor. Besides the continuous automatic decomposition, the user can also trigger the analysis manually. Mapped to a foot pedal, this allows the instrumentalist to trigger a new decomposition only when the sonic material changes, keeping the spatialization otherwise static. Sub-patchers from Cycling 74's Audio Routes package (Cycling74, 2024) are used to allow the user to route the separate spectral streams freely to any track or external output.

3.4 Audio Examples

For the provided audio examples, the file name suffix specifies the channel count of the audio file. Files with the suffix *-1ch* are mono files used as the input for the spatialization, *-4ch* is the spatialized output and *-2ch* is a stereo down-mix of the spatialized output.

The sound file *ex-1* demonstrates the spatialization without any control from the performer. The length of the buffers used for the NMF analysis was set to 4 seconds and the cross fade to 2 seconds to achieve continuous gradual movement in the soundscape.

ex-2 demonstrates the use case of the performer controlling the amount of spatialization in real-time. In the files *ex-2-2ch* and *-4ch* there is no control of the performer with the spatialization set to maximum. This results in audible phasing artifacts when the frequency range of the input is too small and the NMF algorithm decomposes the audio into spectral streams that are very narrow and close to each other, for example if the saxophone plays a slow melody. By controlling the spread of the spatialization (files *ex-2-ctrl-2ch* and *-4ch*), the performer can keep the sound centered for passages with a narrow frequency spectrum and spread the sound for passages that cover a broader range of frequencies.

In both examples, the listener can clearly hear the difference between the input and the spatialized material. By generating continuous movement in the music, the spatialization enables the listener to perceive the different parts of

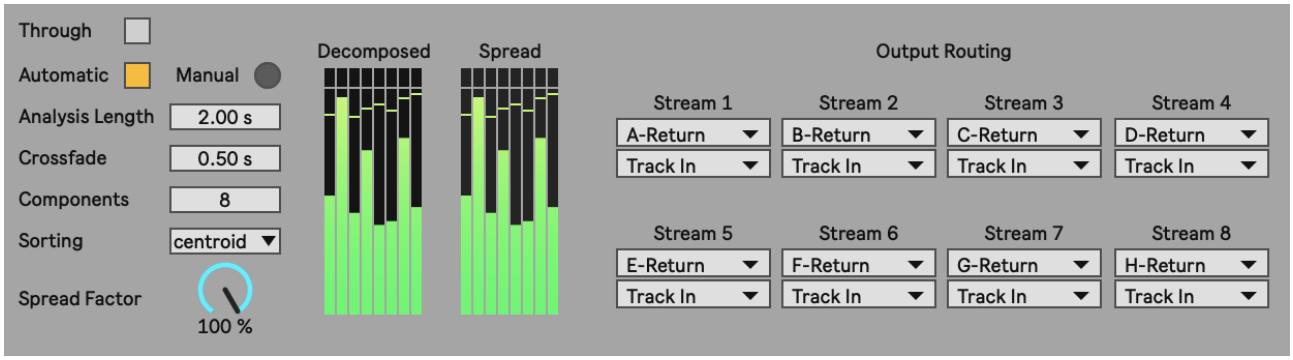

Figure 5: Max for Live implementation of the system.

the saxophone sound as related but distinct sound sources that move through the space and gives depth to the texture of the saxophone sound.

3.5 Evaluation

The proposed system allows the artist to automatically spatialize a mono audio input into a circumspectral soundscape in real-time, while maintaining control over the ordering and amount of the spatialization. Because of the connection between the spectral properties of the sound and the resulting positions and movements in the soundscape, the spatialization becomes a part of the instrument from the perspective of the improviser and allows them to expand the range of their expressive possibilities. Since the system does not need to be trained in advance and the performer is not limited to prepared material, this process can happen in real-time during practice as well as performance.

The flexibility of the channel-based setup allows the system to be used with unconventional speaker setups and to quickly be adapted to changes in the speaker setup, which is essential for a system aimed at live performances. Even if the output of the system is mixed down to stereo, the positions and movements it generates still translate to the listener. This can also be essential to allow the performer to develop new material in their daily practice, without access to multi-speaker setups.

One drawback of the NMF-based solution is the indeterminacy of the audio decomposition. Since one cannot predict the results of the decomposition, it is not possible to set up processing for specific spectral streams that one expects the NMF algorithm to detect. However, spectral analysis can help process objects with different properties differently, even though this processing only depends on the relative differences between objects. This could be improved by creating a wrapper for the audio effects, that only enables the additional processing if certain thresholds for selected spectral properties are reached.

Since movement in the soundscape emerges from the changing decomposition, it is also indeterminate. Currently, it is only possible to decide between "less" movement, by sorting the objects based on their spectral properties, or "more" movement, by not sorting the objects. This could be improved by implementing a more advanced sorting system based on comparison between the current and previous NMF analysis or introducing movement by rotating and mirroring either the bases buffer or the audio channels in the output.

4 APPLICATION TO SOLO BASS-SAXOPHONE

The NMF-based system was used for multiple performances during the course of the research. Here, we focus on one application: an improvised bass-saxophone solo performance by the first author in a recording studio with an octophonic speaker setup. This project showcases the artistic possibilities of the system in its current form, since it does not apply any other forms of spatialization and contains a large range of sonic material.

4.1 Technical Setup

The eight speakers were not identical, but made up of four different stereo pairs and the available space did not allow for a perfectly symmetrical placement of the speakers. The space was also relatively small, which means the bass-saxophone as an acoustic sound source acts as an additional 9th sound source. The setup was therefore chosen in way that integrates the performer into the multi-speaker setup as a front center sound source (Figure 6). This setup is a good example for the adaptability of the system.

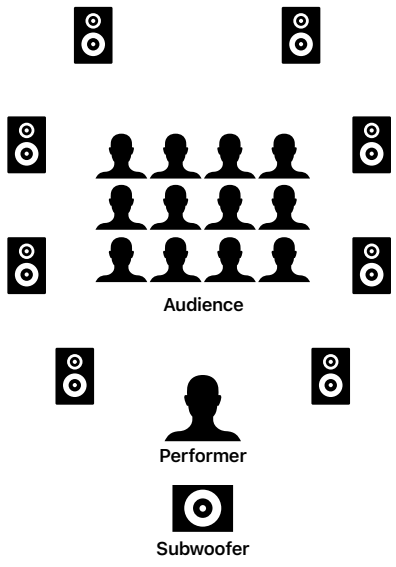

Figure 6: Octophonic setup for solo performance.

The saxophone was recorded using an IntraMic (Levin, 2019) fitted at the opening of the neck of the saxophone inside of the mouthpiece for the main sound, a clip micro-

phone at the bell of the saxophone for the low end of the sound, and a contact microphone with pre-amplifier at the lower body of the saxophone to pick up the key clicks as an additional percussive voice.

In the Max patch, all inputs were summed together and split into two frequency bands with the crossover set to 200 Hz. The lower band was sent directly to the subwoofer, the upper band were decomposed into eight spectral streams. The size of the ring-buffer was set to 4000 ms, which means new bases were generated every 2000 ms. The detected spectral streams were sorted by spectral centroid from low to high and the cross fade between each analysis was set to 500 ms. The two spectral streams with the most low frequency content were mapped to the front speakers, close to the subwoofer. Spectral streams with the highest frequency content were mapped to the back speakers, the furthest away from the saxophone and the subwoofer. By deriving the spatial separation of the spectral streams from their spectral differences, simultaneously occurring spectral streams are more likely to be perceived as separate sounds, especially if there is no simple harmonic ratio between the high frequency content in the back and the low content in the front and if fluctuations in their amplitudes are not synchronized (Bregman and Woszczyk, 2004).

The general amount of spatialization was controlled by the performer using an expression pedal. For material where the frequency range was too small for it to be decomposed, the performing author kept the spread factor at 0 and no spatialization is taking place. As the material became more complex and covered a wider frequency range, the spread factor was increased, until the spectral streams were distributed across all eight channels (Figure 7).

The eight audio channels of the performance were recorded separately. To create the *performance* audio example accompanying this section, the recording was mixed to stereo, panning the front channels 25%, the front-side channels 50%, the rear-side channels 75%, and the rear channels 100% to the left and right, respectively.

4.2 Analysis of the Spatialization

Since the system's behavior is shaped by the specific performer's technique and aesthetic choices, the analysis presented follows a practice-based methodology, treating the performing author's own real-time engagement with the system in a performance context as part of the knowledge production (Pelinski et al., 2025), while acknowledging this remains a single-subject, first-person account.

4.2.1 Multiphonics

The NMF algorithm decomposes the multiphonics into eight small sets of partials, distributed across the speakers. The performer perceived this distribution as emphasizing the beating between closely but not exactly aligned frequencies, that naturally occurs in saxophone multiphonics, with partials seeming to move between speakers.

The algorithm captures dense multiphonics as one spectral stream, while multiphonics with more widely spread partials are decomposed into several separate streams and distributed across multiple speakers. Correspondingly, denser multiphonics were experienced as having a more distinct, stable position, and more spread-out multiphonics as more spread across the space, introducing a layer of perceived movement in the soundscape, with the spatialization as a whole seeming to expand and contract between positions.

4.2.2 Key noises

Since the resonances and mechanical noises differ based on the key, the NMF algorithm decomposes the different clicks into different spectral streams which are distributed across multiple speakers. With each new analysis, the specific spectral streams and their assignment to the speakers changes, which was experienced as key clicks moving through the space over time.

4.2.3 Melodic material

For slower melodic material combined with key clicks, the NMF algorithm does not decompose the notes themselves, but recognizes the notes and the clicks as separate spectral streams which are distributed across the speakers. The performing author experienced melodic material being placed in the space separately from the rhythmic material, appearing to move slowly through the space.

Faster melodic material merges with the key noises of the saxophone. The algorithm detects multiple shorter notes as well as the key clicks as separate spectral streams distributed across the speakers, which was experienced by the performing author as a diffused soundscape without any directional movement of the elements.

4.2.4 Drone

When low, resonant drones are played, the frequency range is wide enough for the algorithm to decompose the sound into multiple components. As a result, phasing did not appear to be an issue. The low end is routed continuously to the subwoofer and front speakers, while upper partials and breath sounds are placed in the rear speakers. The performing author perceived changes in overtones due to voicing as introducing movement both within the sound itself and in its spatial position.

4.2.5 Overtones

Shifts between different overtone voicings function acoustically similar to a frequency sweep. The algorithm decomposes the sound into the different overtones and assigns those to different speakers, which lead to the performing author experiencing the movement between frequencies as movement in the soundscape, with the sound appearing to start in the front and move toward the back as the frequency rose. Rhythmical changes in pitch and amplitude, produced through pitch-bending and air pressure, were perceived as causing the whole drone to move back and forth slightly in the space.

5 CONCLUSION

The goal of this research was to automate real-time audio spatialization to allow improvisers to expand their expressive possibilities. To this end, a system for real-time audio decomposition was implemented, which showed that it is possible to infer spectral streams in real time and assign each stream to its own speaker. By distributing the decomposed audio across the circumspace, the system adds a new spatial dimension to the improvised material based on its spectral properties.

The application in a solo saxophone performance shows how real-time audio decomposition and spatialization can expand a solo improviser's expressive possibilities on an acoustic instrument without requiring additional effects.

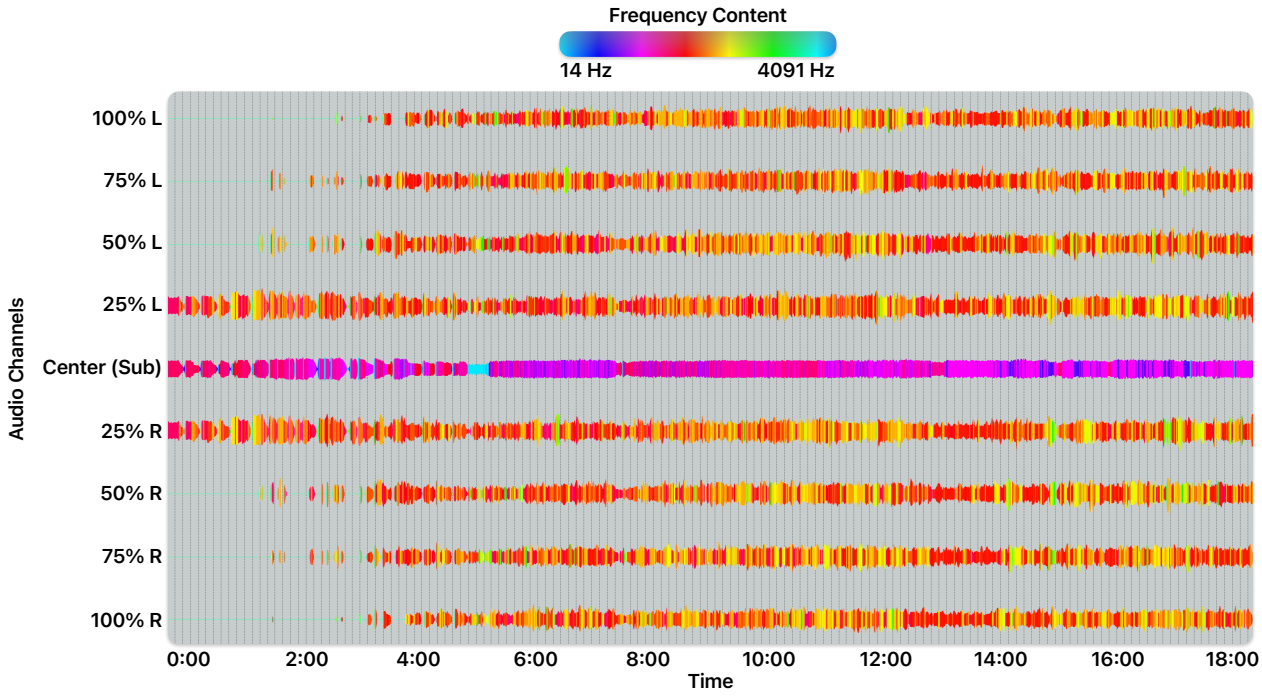

Figure 7: Distribution of the input across the audio channels throughout the performance, with the color representing the frequency content shown as "Spectral Peaks" in Reaper (Cockos Inc., 2026).

Across the examined material, the spatialization highlighted the character of individual sounds, introduced layers of movement, and emphasized the juxtaposition of contrasting sounds, with a minimal amount of manual control.

6 FUTURE WORK

A natural next step for this work is to corroborate the single-subject, first-person account provided in the analysis of the spatialization through listening studies with external participants, which allow the observations reported here to be compared against independent listener evaluations.

Currently, the system only supports mono inputs. Traditionally, acousmatic sound spatialization is done based on a stereo file, since this allows the composer to express more spatiality in the source material. While an acoustic instrument is usually considered a mono sound source, live electronics artists often work with stereo effects. In this case, the stereo sound source has to be summed to mono, losing any panning information, before it can be spatialized using the proposed NMF-based system. There is research that introduces NMF for multichannel audio source separation (Ozerov et al., 2018). Future work could aim to implement this for a real-time spatialization system, in which the spatiality of stereo- and multichannel material could be taken into account in the decomposition process.

The re-synthesis of the detected spectral objects can be used to analyze them and use the results of that analysis to determine their placement in the space. The implementation of a wrapper object for arbitrary audio analysis abstractions could enable the user to provide their own analysis algorithm (like machine learning-based audio classification) to be run on the re-synthesis. This could not only be used to determine the spatial placement of the spectral objects, but also to process the outputs based on their properties.

For improvisers who want more control over the amount of movement, an additional group of Max abstractions could be implemented, based on rotating/switching the position of the bases in the bases buffer after it has been sorted according to the analysis described above. The number of channels to rotate could be controlled in real-time via an expression pedal or a midi controller. If the channels containing the bases are rotated before crossfading from the old bases, this introduces additional movement into the circumspectral soundscape. These additional abstractions could be set up to allow for rotational as well as lateral movement.

7 CODE AND AUDIO AVAILABILITY

All of the scripts, Max patches, and audio files referred to in this research are available in the software and audio folders or at https://gitlab.com/michaelmariaott/nmf-decomposer.

REFERENCES

Austin, L. and Smalley, D. (2000). Sound Diffusion in Composition and Performance: An Interview with Denis Smalley. *Computer Music Journal*, 24(2):10–21. doi: 10.1162/014892600559272.

Boutsidis, C. and Gallopoulos, E. (2008). SVD based initialization: A head start for nonnegative matrix factorization. *Pattern Recognition*, 41(4):1350–1362. doi: 10.1016/j.patcog.2007.09.010.

Bregman, A. and Woszczyk, W. (2004). Controlling the perceptual organization of sound: Guidelines derived

from principles of auditory scene analysis (asa). *Audio Anecdotes: Tools, Tips, and Techniques for Digital Audio*, pages 33–61.

Cannon, J. and Favilla, S. (2010). Expression and Spatial Motion: Playable Ambisonics. *Proceedings of the 2010 Conference on New Interfaces for Musical Expression*, pages 120–124.

Catena, S. and Frisk, H. (2024). A speaker agnostic approach to spatialisation in electroacoustic music. In *Proceedings of the 21st Sound and Music Computing Conference*.

Cockos Inc. (2026). Reaper 7.78. https://www.reaper.fm/. Accessed: 14.08.2026.

Cycling74 (2024). Audio routes 1.5.1. https://maxforlive.com/library/device/5830/audio-routes. Accessed: 01.08.2026.

Cycling74 (2026). Max 9. https://cycling74.com/products/max. Accessed: 01.08.2026.

Févotte, C., Bertin, N., and Durrieu, J.-L. (2009). Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis. *Neural Computation*, 21(3):793–830. doi: 10.1162/neco.2008.04-08-771.

Févotte, C., Vincent, E., and Ozerov, A. (2018). Single-Channel Audio Source Separation with NMF: Divergences, Constraints and Algorithms. In Makino, S., editor, *Audio Source Separation*, pages 1–24. Springer International Publishing, Cham. doi: 10.1007/978-3-319-73031-8_1.

Frank, M., Zotter, F., and Sontacchi, A. (2015). Producing 3D Audio in Ambisonics. *Audio Engineering Society Conference: 57th International Conference: The Future of Audio Entertainment Technology–Cinema, Television and the Internet*.

Khosravi, P. (2011). Circumspectral Sound Diffusion with Csound. *Csound Journal*, 15.

Lee, D. and Seung, H. S. (2000). Algorithms for non-negative matrix factorization. In Leen, T., Dietterich, T., and Tresp, V., editors, *Advances in Neural Information Processing Systems*, volume 13. MIT Press.

Levin, D. (2019). The intramic: Breaking down the revolutionary wind instrument microphone. https://www.horn-fx.com/intramic-article. Accessed: 25.07.2026.

Lossius, T. and Baltazar, P. (2009). Dbap - Distance-Based Amplitude Panning. *Proceedings of the International Computer Music Conference*, pages 489 – 492.

Ozerov, A., Févotte, C., and Vincent, E. (2018). An Introduction to Multichannel NMF for Audio Source Separation. In Makino, S., editor, *Audio Source Separation*, pages 73–94. Springer International Publishing, Cham. doi: 10.1007/978-3-319-73031-8_4.

Pelinski, T., McPherson, A., and Fiebrink, R. (2025). Ways of knowing, ways of writing: technical practice research in new musical instrument design. *Journal of New Music Research*, 53:79–92.

Petermann, D., Wichern, G., Wang, Z.-Q., and Roux, J. L. (2022). The cocktail fork problem: Three-stem audio separation for real-world soundtracks. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 526–530. doi: 10.1109/ICASSP43922.2022.9746005.

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

Pulkki, V. (1997). Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *Journal of the audio engineering society*, 45:456–466.

Roma, Gerard, Green, Owen, and Tremblay, Pierre Alexandre (2020). Audio Morphing Using Matrix Decomposition and Optimal Transport. *Proceedings of the 23rd International Conference on Digital Audio Effects*, 1:147–154.

Sarkar, S. (2024). *Time-Domain Music Source Separation for Choirs and Ensembles*. PhD thesis, Queen Mary University of London School of Electronic Engineering and Computer Science.

Smalley, D. (1997). Spectromorphology: Explaining sound-shapes. *Organised Sound*, 2(2):107–126. doi: 10.1017/S1355771897009059.

Smalley, D. (2007). Space-form and the acousmatic image. *Organised Sound*, 12(1):35–58. doi: 10.1017/S1355771807001665.

The FluCoMa Project (2022a). BufNMFSeed. https://learn.flucoma.org/reference/bufnmfseed/. Accessed: 01.08.2026.

The FluCoMa Project (2022b). NMFFilter. https://learn.flucoma.org/reference/nmffilter/. Accessed: 01.08.2026.

Tremblay, P. A., Green, O., Roma, G., Bradbury, J., Moore, T., Hart, J., and Harker, A. (2022). The Fluid Corpus Manipulation Toolbox. Zenodo. doi: /10.5281/zenodo.6834643.

Watcharasupat, K. N. and Lerch, A. (2024). A stem-agnostic single-decoder system for music source separation beyond four stems. *Proceedings of the 25th International Society for Music Information Retrieval Conference*, pages 1051–1059.

Watcharasupat, K. N., Wu, C.-W., Ding, Y., Orife, I., Hipple, A. J., Williams, P. A., Kramer, S., Lerch, A., and Wolcott, W. (2024). A generalized bandsplit neural network for cinematic audio source separation. *IEEE Open Journal of Signal Processing*, 5:73–81. doi: 10.1109/ojsp.2023.3339428.

Yun-Ning, Hung, Pereira, I., and Korzeniowski, F. (2025). Moises-light: Resource-efficient band-split u-net for music source separation. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*.