

Adaptive Real-Time Spatialization for Live Saxophone 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

This performance explores the transformation of the solo bass saxophone into a spatial soundscape through real-time audio decomposition. Using a custom Max system based on Non-Negative Matrix Factorization (NMF), the sound of the saxophone is dynamically decomposed into separate spectral streams, which are then mapped to a multi-speaker setup based on their frequency content. Monophonic melodies remain localized, while multiphonics expand to envelop the audience, with partials drifting between speakers and the saxophone’s key clicks dispersing as independent percussive voices. The performer controls the spatialization with an expression pedal, enabling fluid transitions between localized sounds and immersive, multichannel textures. This approach allows the performer to use the multi-speaker setup as an extension of the instrument, translating the sound of the improvisation into spatial placement and movement.

1 SPATIALITY AND SPATIALIZATION

All music can imply spatiality (Austin and Smalley, 2000), both in its spectral content (Smalley, 1997), as well as in the placements and trajectories implied by the development of the sound spectra over time (Smalley, 2007). Spatialization can enhance the experience of that implied spatiality by translating it into the physical space.

In live performance settings, spatialization usually means placing and moving sources using either channel-based or coordinate-based systems (Peters et al., 2011; Frank et al., 2015; Pulkki, 1997; Lossius and Baltazar, 2009). Not only is spatial placement usually considered separately from the instrumental practice in these approaches due to the large number of parameters to control (Austin and Smalley, 2000), but spatializing a single monophonic sound-source additionally requires real-time decomposition of the sound.

2 PERFORMANCE CONCEPT

In this improvised saxophone performance, machine learning is leveraged to allow the performer to control the spatialization as part of the instrument. The spatialization is controlled through the performers instrumental technique and hardware controls, where an expression pedal controls the size of the soundscape.

Using a multi-speaker setup around the audience, the decomposition-based spatialization turns the saxophone from a single sound source into a soundscape enveloping the audience. The partials of multiphonics are spread around the listener and key clicks are dispersed based on their resonant frequencies. Movement in the soundscape is generated both by changes in the results of the reoccurring indeterministic analysis, as well as changes in the musical material. This allows the performer to develop an improvisational vocabulary on the saxophone that incorporates the spatial placement as part of the sound itself, rather than a separately controlled process.

The performance explores these new expressive possibilities through both traditional and extended techniques on the bass saxophone.

3 SYSTEM DESIGN

To translate the sound of a single monophonic sound source to a soundscape, which envelops the listener (Smalley, 1997), we developed a system to decompose the audio of a single monophonic source in real-time. The system design and architecture are described in more detail in the companion full paper submission (Ott and Damsma, 2026).

Audio decomposition is performed with Non-Negative Matrix Factorization (NMF). The NMF algorithm for matrix decomposition (Lee and Seung, 1999) can be applied to audio to decompose an audio file into spectral streams (Févotte et al., 2009, 2018). We used the NMF implementation of the FluCoMa (Tremblay et al., 2022) package for Max (Cycling74, 2026) to implement a real-time audio decomposition system (Figure 1).

The audio is recorded into a buffer and analyzed on a separate processing thread. On the main thread, the incoming audio is decomposed into spectral streams in real-time based on the results of the analysis. Every analysis triggers a crossfade from the previous decomposition to the new decomposition. The results of each decomposition are indeterministic, but we implemented a sorting algorithm, which allows the spectral streams to be assigned to channels based on properties like spectral centroid or kurtosis. The performer can control the decomposition by setting the length of the analysis buffer, the number of components the audio is decomposed into, the sorting algorithm and the spread, which determines the size of the resulting soundscape by controlling how far the individual components are spread from the center of the sound stage.

*www.michaelmariaott.com

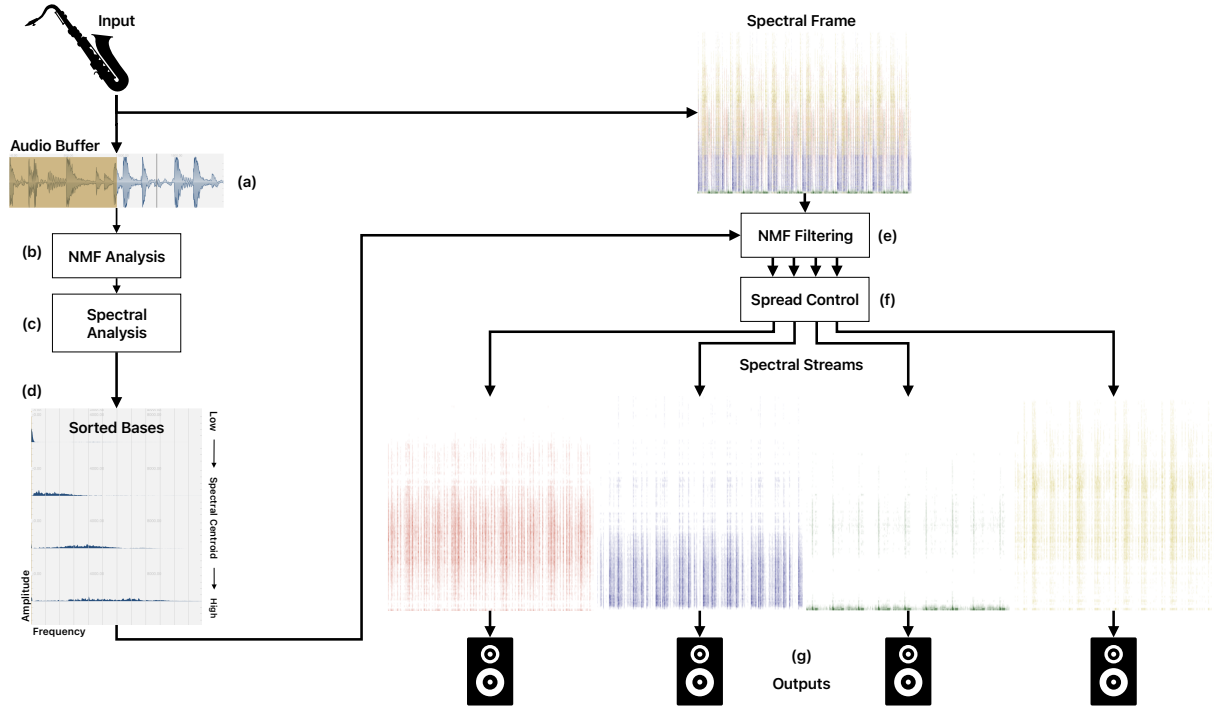


Figure 1: The audio processing pipeline: Recording the input (a), analyzing (b - d) and filtering (e) it using NMF, controlling the spread (f) and sending the spectral streams to separate outputs (g).

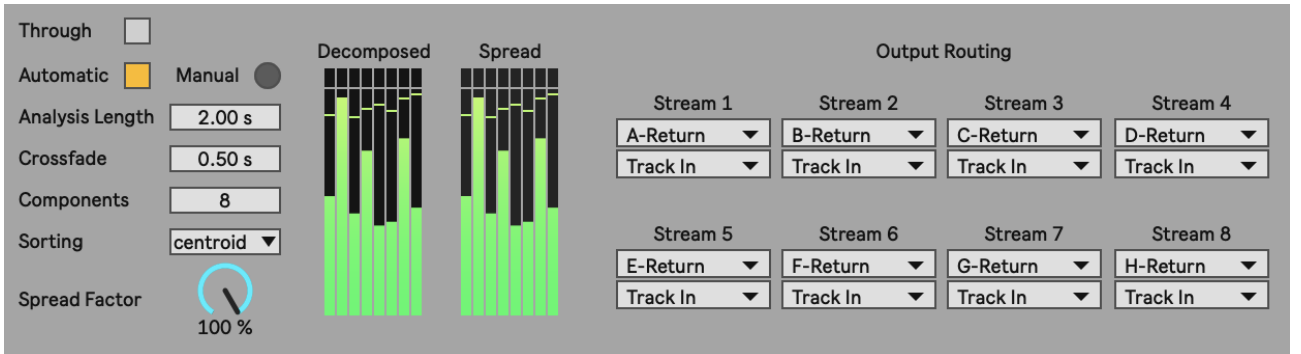


Figure 2: Max for Live implementation.

The Max for Live device (Figure 2) we implemented allows for all settings to be controlled during performance. In practice, however, it is enough to control only the spread, since the effect of the spatialization is determined mainly by the sound of the instrument. This allows the performer to control size and movement of the spatialization without requiring too much cognitive overhead.

4 IMPROVISING WITH THE SYSTEM

The saxophone sound is decomposed differently for different playing techniques. The performer exploits this feature, creating changing spatial textures by varying their sound.

If the improviser plays only a single note, the audio gets decomposed into components that are similar enough to each other to cause phasing effects. This can be used deliberately as an effect or avoided by keeping the spread low. If the note has enough spectral range (e.g., a low note with prominent overtones), different parts of the spectrum

will be assigned to each channel. The sound of the note envelops the listener and changes in timbre, which the performer can control through changes in the voicing or air pressure, result in movement in the soundscape.

During faster melodic passages, where multiple notes are recorded into the analysis buffer, the system will treat each note as a separate spectral stream. Notes or note groups are assigned to different channels, resulting in the melody moving throughout the performance space. By controlling the spread of the spatialization, melodies can spread out and contract.

If the improvisational gestures focus on timbre, like multiphonics or overtone voicings, the sound is decomposed into multiple timbral layers. The harmonic partials are assigned to different speakers and movement emerges from changes in the timbre.

The decomposition of percussive material results in assignment based on resonant frequencies in the material (e.g., slap tongue) or timbral layers (e.g., air noises).

5 REFLECTION

The developed system allows the improviser to draw from their instrumental experience to generate a soundscape enveloping the audience in real-time. The resulting soundscape consistently reflects the spectral characteristics of the source material. This allows the spatialization to become an integral component of the improvisation, extending the performers expressive possibilities during their improvisation.

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.
- Cycling74 (2026). Max. <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*.
- Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791. doi: 10.1038/44565.
- Lossius, T. and Baltazar, P. (2009). Dbap - Distance-Based Amplitude Panning. *Proceedings of the International Computer Music Conference*, pages 489 – 492.
- Ott, M. and Damsma, A. (2026). Adaptive real-time spatialization for live music improvisation. *Proceedings of the 7th Conference on AI Music Creativity (AIMC)*.
- 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.
- Smalley, D. (1997). Spectromorphology: Explaining sound-shapes. *Organised Sound*, 2(2):107–126. doi: 10.1017/S1355771897009059.
- Smalley, D. (2007). Space-form and the acoustic image. *Organised Sound*, 12(1):35–58. doi: 10.1017/S1355771807001665.
- 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.