

## Analyse durch Synthese: Musikverarbeitung von Blasmusikaufnahmen

Stefan Balke, Vlora Arifi-Müller, Simon Schwär, Meinard Müller

International Audio Laboratories Erlangen, Deutschland, E-Mail: stefan.balke@audiolabs-erlangen.de

### Einleitung

Blasorchester sind ein zentraler Bestandteil der Amateurmusikultur im deutschsprachigen Raum, im Forschungsfeld des Music Information Retrieval (MIR) bislang jedoch deutlich unterrepräsentiert. Insbesondere fehlen systematische Analysen und datenbasierte Ansätze für Blasinstrumente in komplexen Ensembleszenarien, etwa in den Bereichen Transkription [3], Quellentrennung [1] und Klanganalyse [16].

Blasmusik stellt spezifische Anforderungen an MIR-Methoden. Klang und Intonation entstehen durch kontinuierliche Atemführung, flexible Artikulation und komplexe Obertonstrukturen. In größeren Besetzungen kommen klangliche Überlagerungen, räumliche Effekte sowie dynamische Wechsel hinzu, die bestehende Modelle vor erhebliche Herausforderungen stellen. Zwar existieren zahlreiche Arbeiten zur Klangerzeugung von Blasinstrumenten, unter anderem im Kontext physikalischer Modellierung [8, 11, 17] und differenzierbarer Syntheseverfahren [4, 5, 15], jedoch fehlen bislang integrative Ansätze, die Analyse und Synthese systematisch miteinander verbinden. Ein vielversprechender Rahmen ist Differentiable Digital Signal Processing (DDSP) [6, 10], das klassische signalverarbeitende Modelle mit lernbasierten Verfahren kombiniert und dadurch interpretierbare Parameter zugänglich macht.

In diesem Beitrag verfolgen wir einen Analyse-durch-Synthese-Ansatz, bei dem musikalische Steuerparameter wie Tonhöhe, Dynamik, Klangfarbe und Artikulation aus Audio- und Notendaten extrahiert und zur Rekonstruktion des Signals genutzt werden. Zum Einsatz kommen insbesondere Pulsetable-Synthese [14] sowie DDSP-Modelle [6, 10]. Der Vergleich zwischen synthetischen und realen Aufnahmen ermöglicht es, Analyseverfahren gezielt zu verfeinern und modellierte Parameter musikalisch zu interpretieren.

Als Anwendungsszenario dienen die Suiten für Blasmusik von Gustav Holst (siehe Abbildung 1), die aufgrund ihrer orchestralen Struktur und klanglichen Vielfalt besonders geeignet sind. Aufbauend auf Einzelstimmen, Mischungen und Notentexten entsteht ein multimodaler Datensatz, der an bestehende Arbeiten wie ChoraleBricks [2] anknüpft und neue notentext-informierte Werkzeuge für Analyse, Lehre und künstlerische Anwendungen ermöglicht.

### Analyse durch Synthese

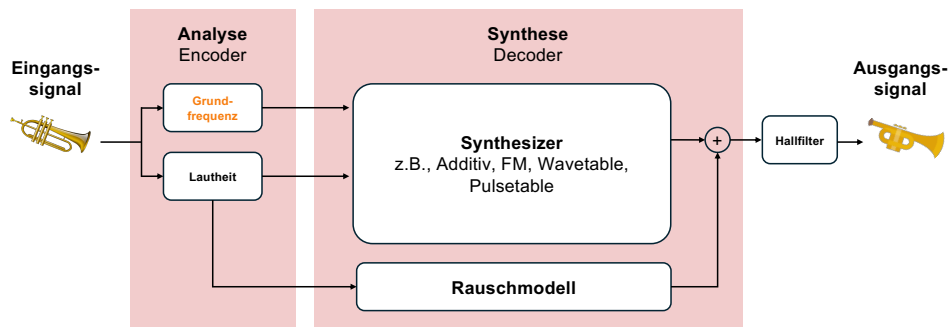
Ein zentraler Bestandteil unseres Ansatzes ist die Analyse durch Synthese, bei der ein Audiosignal durch explizit modellierte und interpretierbare Parameter beschrieben und rekonstruiert wird. Dieses Paradigma wurde ursprünglich im Kontext der Sprachverarbeitung eingeführt [9] und

**Abbildung 1:** Auszug aus der Partitur von Gustav Holsts First Suite in E-flat, 1. Satz. Die Passage veranschaulicht die dichte instrumentale Satzstruktur und die klanglichen Überlagerungen, die für Blasorchester charakteristisch sind und MIR-Verfahren vor besondere Herausforderungen stellen.

lässt sich in natürlicher Weise auf musikalische Signale übertragen, bei denen klanglich relevante Eigenschaften nicht nur erkannt, sondern gezielt modelliert und variiert werden sollen. Ziel ist es, musikalisch aussagekräftige Merkmale direkt in der Signalrepräsentation abzubilden und so eine Brücke zwischen signalverarbeitenden und musikalischen Beschreibungen zu schlagen.

Als methodischer Rahmen dient Differentiable Digital Signal Processing (DDSP) [6, 10]. Hierbei werden klassische Bausteine der digitalen Signalverarbeitung, etwa Oszillatoren, Filter oder Hüllkurven, in differenzierbare Modelle integriert. Aufbauend auf Arbeiten zur neuronalen Audiosynthese [7, 12] sowie zu differenzierbaren Syntheseverfahren wie FM-Synthese [4], modalen physikalischen Modellen [11] oder Wavetable-Synthese [15], ermöglicht DDSP die Kombination von Domänenwissen mit datengetriebenen Lernverfahren. Dies führt zu effizienten und zugleich interpretierbaren Repräsentationen, die eine gezielte Kontrolle über das erzeugte Klangsignal erlauben. Im Vergleich zu rein datengetriebenen Ansätzen reduziert sich der Bedarf an Trainingsdaten, während die Nachvollziehbarkeit der Modellparameter erhalten bleibt.

Für Blasinstrumente ist dieser Ansatz besonders geeignet,

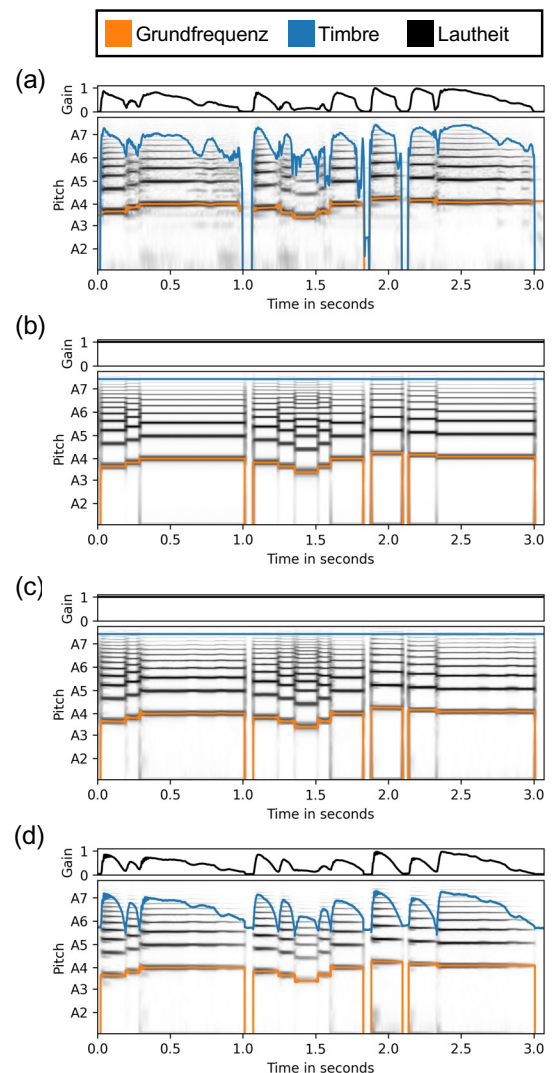


**Abbildung 2:** Schematische Darstellung der verwendeten Analyse-Synthese-Verarbeitungskette. In der Analyseschicht werden Steuerparameter wie Grundfrequenz und Lautheit aus dem Eingangssignal extrahiert und an die Syntheseschicht übergeben. Dort werden die harmonischen Anteile mit einem tonalen Synthesizer und die nicht-tonalen Anteile mit einem Rauschmodell rekonstruiert. Eine abschließende Faltung mit einem Hallfilter dient der Annäherung an reale Aufnahmebedingungen.

da deren Klang wesentlich durch kontinuierliche Steuergrößen wie Tonhöhe, Lautheit, Artikulation und spektrale Formung geprägt ist. In unserem Ansatz steht insbesondere die Pulsetable-Synthese im Mittelpunkt [13, 14]. Dieses Verfahren basiert auf der periodischen Wiederholung kurzer Wellenformprototypen und ist für bestimmte Blasinstrumente besonders geeignet, da es dem physikalischen Anregungsmechanismus von Rohrblatt- und Blechblasinstrumenten nahekommmt. Im Unterschied zur Wavetable-Synthese werden die Pulssignale dabei nicht auf die jeweilige Periodenlänge skaliert, sondern in ihrer Form beibehalten, durch Zero-Padding erweitert, und mit variabler Wiederholrate eingesetzt. Dadurch eignet sich der Ansatz insbesondere für Klänge, deren spektrale Hüllkurve über verschiedene Tonhöhen hinweg weitgehend stabil bleibt [14].

Abbildung 2 zeigt die grundlegende Analyse-Synthese-Verarbeitungskette. In der Analyseschicht werden zunächst Steuerparameter wie Grundfrequenz und Lautheit aus dem Eingangssignal extrahiert und an die Syntheseschicht übergeben. Diese besteht aus einem tonalen Synthesizer zur Modellierung der harmonischen Komponenten sowie einem Rauschmodell zur Erfassung nicht-tonaler Anteile, etwa von Einschwingvorgängen oder Atemgeräuschen. Eine abschließende Faltung mit einem Hallfilter dient der Annäherung an reale Aufnahmebedingungen. Der Vergleich zwischen Eingangs- und Ausgangssignal ermöglicht eine iterative Optimierung der Parameter, beispielsweise mittels Gradientenabstieg.

Die Rolle der einzelnen Steuerparameter wird anhand eines Trompetensignals in Abbildung 3 veranschaulicht. Eine Rekonstruktion mit konstanten Parametern erhält zwar grundlegende Obertonstrukturen, verliert jedoch wesentliche dynamische und expressive Eigenschaften (siehe (b)). Die schrittweise Wiedereinführung von Grundfrequenzvariationen, Timbre und Lautheit führt zu einer deutlich verbesserten Annäherung an das Originalsignal, sowohl im Spektrogramm als auch im Klangeindruck (siehe (c)). Insbesondere feine Intonationsschwankungen, klangfarbliche Veränderungen und dynamische Verläufe werden besser erfasst (siehe (d)). Diese Beobachtungen zeigen, dass bereits wenige, musikalisch interpretierbare Parameter zentrale Aspekte des Klangcharakters von Blasin-



**Abbildung 3:** Einfluss zentraler Steuerparameter auf die Rekonstruktion eines Trompetensignals. (a) Originalspektrogramm mit den zeitlichen Verläufen von Grundfrequenz, Timbre und Lautheit. (b) Rekonstruktion mit konstanten Werten für Timbre und Lautheit. (c) Rekonstruktion unter zusätzlicher Berücksichtigung von Grundfrequenzschwankungen. (d) Rekonstruktion mit zeitvariierenden Verläufen aller Steuerparameter. Die Abbildung verdeutlicht, wie die sukzessive Einbeziehung der Parameter zu einer verbesserten Annäherung an das Originalsignal führt.



**Abbildung 4:** Multimodale Datenerfassung im ChoraleBricks-Datensatz [2]. (a) Anfang des vierstimmigen Chorals “Jesu geh voran” als musikalische Grundlage in SATB-Struktur. (b) Aufnahmeprozess mit Dirigiervideo zur Synchronisation der einzelnen Stimmen. (c) Beispielhafte Annotationen, einschließlich Grundfrequenzverläufen sowie Annotationen von Notenanfangs- und -endzeiten.

strumenten modellieren können. Ähnliche Befunde zur Bedeutung strukturierter und kontrollierbarer Parameterisierungen wurden auch in verwandten DDSP-Arbeiten berichtet [6, 10, 14].

Insgesamt fungiert der Syntheseprozess nicht nur als Mittel zur Signalrekonstruktion, sondern zugleich als Analyserwerkzeug. Die kontrollierte Rekonstruktion macht hör- und sichtbar, welche Parameter für den Klangcharakter entscheidend sind und welche Aspekte durch das Modell nur unzureichend erfasst werden. Damit eröffnet der Analyse-durch-Synthese-Ansatz neue Perspektiven für das Verständnis musikalischer Ausdrucksparameter sowie für notentext-informierte Anwendungen in Analyse, Lehre und künstlerischer Praxis.

## BlasmusikszENARIO: Datengrundlage und Perspektive

Die Übertragung des beschriebenen Analyse-durch-Synthese-Ansatzes auf realistische Ensembleszenarien erfordert eine geeignete multimodale Datengrundlage. Insbesondere im Bereich der Blasmusik sind entsprechende Datensätze bislang nur begrenzt verfügbar, vor allem wenn kontrollierte Aufnahmebedingungen, präzise zeitliche Synchronisation und detaillierte Annotationen gefordert sind.

Als Ausgangspunkt dient der ChoraleBricks-Datensatz [2], der gezielt für Analyse- und Syntheseaufgaben konzipiert wurde. Die Daten basieren auf vierstimmigen Choralensätzen (SATB), die eine klar strukturierte musikalische Grundlage bieten (vgl. Abbildung 4a). Die einzelnen Stimmen wurden von Musikerinnen und Musikern auf verschiedenen Blasinstrumenten separat und synchron zu einem Dirigiervideo aufgenommen (Abbildung 4b). Dieses Vorgehen ermöglicht eine präzise zeitliche Abstimmung bei gleichzeitiger Vermeidung von Übersprechen. Ergänzend wurden die Aufnahmen mit dem Notentext abgeglichen und um Annotationen wie Grundfrequenzverläufe sowie Notenanfangs- und -endzeiten erweitert (Abbildung 4c). Der resultierende Datensatz verknüpft mehrere Modalitäten und erlaubt systematische Experimente unter kontrollierten Bedingungen.

Aufbauend auf diesen kontrollierten Szenarien wird der Ansatz im nächsten Schritt auf komplexere Ensemblesituationen erweitert. Hierzu werden die First Suite (1909) und



**Abbildung 5:** Beispielhafte Orchesterbesetzung und Sitzordnung für Holsts First Suite mit 19 Blasinstrumenten und Schlagwerk. Die Darstellung illustriert die instrumentale Vielfalt und räumliche Anordnung als Grundlage komplexer Ensembleszenarien.

Second Suite (1911) von Gustav Holst herangezogen, die zum Standardrepertoire sinfonischer Blasorchester zählen. Der reduzierte Partiturauszug in Abbildung 1 verdeutlicht die dichte Satzstruktur und die charakteristischen klanglichen Überlagerungen. Ergänzend zeigt Abbildung 5 eine typische Orchesterbesetzung und Sitzordnung, die die instrumentale Vielfalt sowie die räumliche Komplexität solcher Ensembleszenarien illustriert. Diese Eigenschaften machen die Werke zu einem anspruchsvollen Testfeld für Analyse- und Syntheseverfahren im orchestralen Kontext.

Die geplanten Aufnahmen orientieren sich methodisch an ChoraleBricks, werden jedoch um zentrale Aspekte erweitert. Neben einer Referenzaufnahme eines vollständigen Blasorchesters werden die Einzelstimmen separat eingespielt, um eine detaillierte und kontrollierte Datengrundlage zu schaffen. Zusätzlich wird das Dirigat videografisch erfasst, sodass neben Audio- und Notendaten weitere Modalitäten für die Analyse verfügbar werden. Dieses Aufnahme-konzept wird in einem musikalisch realistischen Probenkontext umgesetzt, in dem Musikerinnen und Musiker die Werke gemeinsam erarbeiten und der Probenprozess durch Audio- und Videoaufnahmen dokumentiert wird. Dadurch entstehen nicht nur Daten für die Signalverarbeitung, sondern auch Einblicke in musikalische Interaktion, Klangbalance und Zusammenspiel im Ensemble.

Durch die Kombination aus kontrollierten Aufnahmen,

orchestralen Referenzdaten und praxisnahen Probensituationen entsteht eine Brücke zwischen experimentellen Studien und realen Anwendungen. Dies ermöglicht es, Analyseverfahren unter realistischen Bedingungen zu evaluieren und zugleich neue Perspektiven für den Einsatz datengetriebener Methoden in der musikalischen Praxis zu eröffnen.

## Zusammenfassung und Ausblick

Blasorchester sind im Music Information Retrieval (MIR) bislang unterrepräsentiert, obwohl ihre klangliche Komplexität besondere Herausforderungen für datengetriebene Methoden darstellt. In diesem Beitrag zeigen wir, wie Analyse-durch-Synthese auf Basis von DDSF eine interpretierbare Modellierung von Blasinstrumentklängen ermöglicht. Mit ChoraleBricks steht eine kontrollierte Datengrundlage zur Verfügung, die im nächsten Schritt durch realistische Orchesteraufnahmen der Holst-Suiten erweitert wird. Auf diese Weise schafft der Beitrag eine Grundlage, um Analyse-durch-Synthese-Verfahren systematisch auf komplexe Blasorchesteraufnahmen zu übertragen und für notentextinformierte Anwendungen in Analyse, Lehre und künstlerischer Praxis nutzbar zu machen. Zur nachhaltigen Verankerung von Musik und Wissenschaft arbeiten wir zudem aktuell am Aufbau eines Blasmusikzentrums (BMZ) in Höxter, das als Schnittstelle zwischen Forschung, Ausbildung und musikalischer Praxis dienen soll.

**Danksagung:** Diese Arbeit wurde durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Projekte Nr. 500643750 (MU 2686/15-1) und Nr. 555525568 (MU 2686/18-1) gefördert. Die Aufnahmen der Holst-Suiten werden durch den Landesmusikrat NRW sowie das Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen unterstützt. Die International Audio Laboratories Erlangen sind eine gemeinsame Einrichtung der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) und des Fraunhofer-Instituts für Integrierte Schaltungen IIS.

## Literatur

- [1] Araki, S., Ito, N., Haeb-Umbach, R., Wichern, G., Wang, Z.-Q. & Mitsufuji, Y.: 30+ Years of Source Separation Research: Achievements and Future Challenges. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 1–5 (2025).
- [2] Balke, S., Berndt, A. & Müller, M.: ChoraleBricks: A Modular Multitrack Dataset for Wind Music Research. In: Transaction of the International Society for Music Information Retrieval (TISMIR) (2025), **8**, 1: 39–54.
- [3] Benetos, E., Dixon, S., Duan, Z. & Ewert, S.: Automatic Music Transcription: An Overview. In: IEEE Signal Processing Magazine (2019), **36**, 1: 20–30.
- [4] Caspe, F., McPherson, A. & Sandler, M.: DDX7: Differentiable FM Synthesis of Musical Instrument Sounds. In: Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), 608–616. Bengaluru, India (2022).
- [5] Derenyi, I. & Dannenberg, R. B.: Synthesizing Trumpet Performances. In: Proceedings of the International Computer Music Conference (ICMC). Ann Arbor, Michigan, USA (1998).
- [6] Engel, J., Hantrakul, L., Gu, C. & Roberts, A.: DDSF: Differentiable Digital Signal Processing. In: Proceedings of the International Conference on Learning Representations (ICLR). Virtual (2020).
- [7] Engel, J. H., Resnick, C., Roberts, A., Dieleman, S., Norouzi, M., Eck, D. & Simonyan, K.: Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders. In: Proceedings of the International Conference on Machine Learning (ICML), Bd. 70, 1068–1077. Sydney, Australia (2017).
- [8] Fletcher, N. H. & Rossing, T. D.: The Physics of Musical Instruments. 2nd Aufl. Springer, Berlin, Germany (1998).
- [9] Halle, M. & Stevens, K. N.: Speech recognition: A model and a program for research. In: IRE Transactions on Information Theory (1962), **8**, 2: 155–159.
- [10] Hayes, B., Shier, J., Fazekas, G., McPherson, A. & Saitis, C.: A review of differentiable digital signal processing for music and speech synthesis. In: Frontiers in Signal Processing (2024), **3**.
- [11] Lee, J. W., Park, J., Choi, M. J. & Lee, K.: Differentiable Modal Synthesis for Physical Modeling of Planar String Sound and Motion Simulation. In: Advances in Neural Information Processing Systems (NeurIPS). Vancouver, BC, Canada (2024).
- [12] Maman, B., Zeitler, J., Müller, M. & Bermano, A. H.: Multi-Aspect Conditioning for Diffusion-Based Music Synthesis: Enhancing Realism and Acoustic Control. In: IEEE Transactions on Audio, Speech, and Language Processing (2025), **33**: 68–81.
- [13] Oehler, M. & Reuter, C.: Dynamic Excitation Impulse Modification as a Foundation of a Synthesis and Analysis System for Wind Instrument Sounds. In: Mathematics and Computation in Music (MCM), 189–197. Berlin, Germany (2009).
- [14] Schwär, S., Dittmar, C., Balke, S. & Müller, M.: Differentiable Pulsetable Synthesis for Wind Instrument Modeling. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain (2026).
- [15] Shan, S., Hantrakul, L., Chen, J., Avent, M. & Trevelyan, D.: Differentiable Wavetable Synthesis. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4598–4602. Singapore (2022).
- [16] Siedenburg, K., Saitis, C., McAdams, S., Popper, A. N. & Fay, R. R. [Hrsg.] : Timbre: Acoustics, Perception, and Cognition. Springer Handbook of Auditory Research. Springer (2019).
- [17] Smith, J. O.: Physical modeling using digital waveguides. In: Computer Music Journal (1992), **16**, 4: 74–91.