Teaching Popular Music Analysis with Source Separation

MICHÈLE DUGUAY

Music source separation is a technology that converts a recorded song into multiple files that contain different components of the original track. Typically, source separation tools split a recording into four discrete components: vocals, percussion, bass, and "other." Methods for isolating components of a recording have been proposed and experimented with for decades, with varying degrees of success. In the late 2010s, developers began using neural networks, a form of artificial intelligence, to design source separation methods. Since then, the performance and availability of source separation models has rapidly improved.

In this article, I aim to demystify and contextualize source separation to explore the ways in which it could impact the study of music analysis while encouraging productive discussions on the relationship between music analysis and artificial intelligence. Access to high-quality isolated instrumental tracks has several useful applications in the study of popular music: consider for instance the way an instructor could use the technology to design specific class activities that focus on only a portion of a track. Source separation, moreover, can play a central role in discussions on notation, musical structure, and the role of technology in music analysis. More importantly, it is an ideal tool for illustrating some of the ethical considerations—transparency, bias, and accessibility—that arise when using AI-based technologies.

Introduction

For the past three years, I have taught an upper-level undergraduate course titled "Analyzing Popular Music." The class is a writing-intensive elective with no prerequisites and is open to students with little to no previous musical experience. Although they may not play an instrument or have previously studied music fundamentals, the typical student enrolling in the class is eager to engage analytically with different aspects of popular music. For this reason, I structure the class around timbre, form, flow, and texture. These musical parameters are central to many popular music genres but do not require extensive training in harmony. Rather than relying on different forms of notation—chord charts, staff notation, or lead sheets—the course relies on recordings.

The first week of class focuses on Allan Moore's functional layers (2012). Moore identifies four functional layers in popular music: the beat layer expresses a regular pattern of beats, the bass layer consists of repeated notes or melodies articulated

in a lower register, the melodic layer is typically performed by a vocalist, and the harmonic filler layer completes the texture with chords, melodic patterns, and other types of ornamentation. An understanding of textural layers allows students to develop a common vocabulary for discussion of what they hear, teaches them to interpret the role of each instrument in a recording, and establishes a conceptual framework for eventual discussions of form, meter, and timbre. As a first exercise, I ask students to identify the functional layers in Nina Simone's "My Baby Just Cares for Me" (1959). We listen to the first audio file in Audio Example 1: Simone sings and plays piano, accompanied by a bassist and a drummer. On the first listen, students identify the four instruments.

a) Full Mix b) Vocals c) Drums d) Bass e) Other

The audio files for this example are hosted on the journal website here.

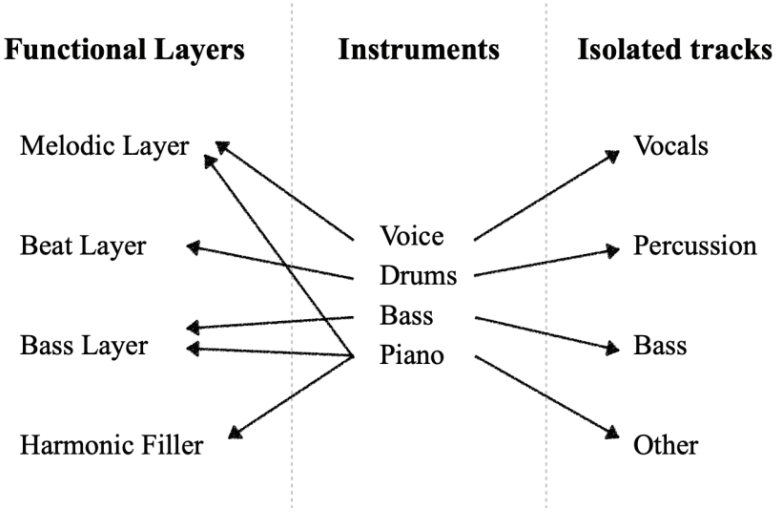
Audio Example 1
Nina Simone, "My Baby Just Cares for Me," 0:00–0:42.

We then listen to the "vocals" file, which contains Simone's voice, the "drums" file, which contains the percussion, the "bass" file, which contains the upright bass, and the "other" file, in which the piano has been placed.[1] We also listen to combinations (bass/percussion, piano/bass, bass/percussion/piano) so that they can hear the different ways in which the instruments interact with one another. For many students used to listening to music casually, identifying instruments and paying attention to textural variations are unfamiliar tasks. These isolated instrumental tracks are a helpful tool to guide students as they practice the focused, deliberate listening necessary for music analysis.

On the second listen of the full track, I ask students to draw on what they just heard to determine which instruments contribute to which functional layer. At first glance, it may seem that each instrument corresponds to one layer: the voice is the melodic layer, the drums are the beat layer, the bass is the bass layer, and the piano is the harmonic filler layer. After some discussion, however, students note that the situation is more complex (Example 1). Although Simone's right hand provides harmonic filler on the piano throughout the track, her left hand participates in the functional bass

1 I have generated these tracks from the full mix, using source separation model Demucs v4 (Rouard et al. 2023). The tracks created through the source separation technology are not perfect. Some components of the percussion, for instance, are still audible in the "vocals" file.

layer because it frequently doubles the bass line. At 1:22, Simone's piano solo takes on the melodic function. Through this exercise, students learn to grasp the difference between an instrument's identity, which is stable, and its function within a song's texture, which is contextual.



Example 1
Distinguishing between functional layers, instrumentation, and isolated instrumental tracks obtained with source separation. Nina Simone, "My Baby Just Cares for Me" (1959).

At this point in the lesson, I can introduce more ambiguous or complex examples. In Audio Example 2, I have separated a portion of Japanese Breakfast's song "Paprika" (2021) into four tracks. With background vocals, strings, and brass, the song has a denser instrumental texture than "My Baby Just Cares for Me." The "vocals" file contains all sung components of the excerpt: Michelle Zauner's reverberated vocals and the backing harmonies. The electric bass and the drum kit are placed in their respective "bass" and "drums" files, while the remaining instruments—synthesizer, brass, and strings—are categorized as "other." Source separation can thus serve as a listening tool for the more interpretive task at hand: the analysis of texture. Consider for instance the trumpet line in the "Other" track. Does it take on a melodic function, replacing the voice at key moments in the chorus? Or does it correspond to Megan Lavengood's "novelty layer" (2020), which complements the melodic layer with marked timbral features that stand out from the rest of the ensemble?

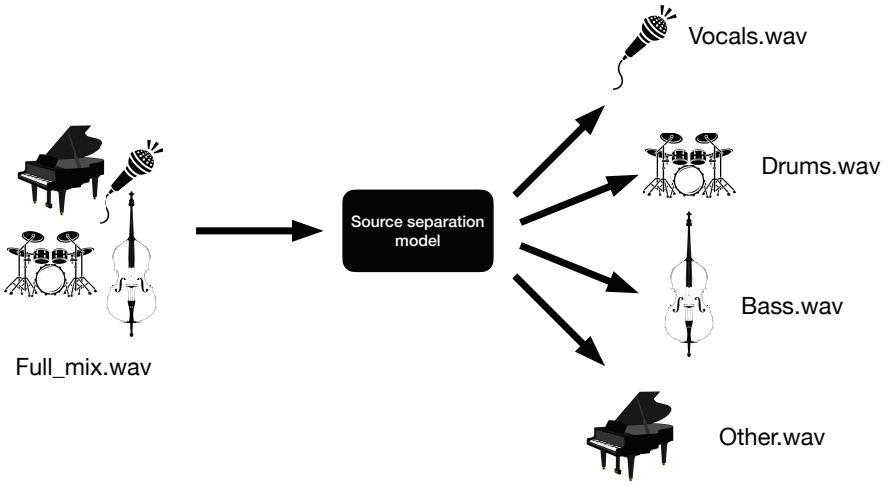
a) Full Mix b) Vocals c) Drums d) Bass e) Other

Audio Example 2
Japanese Breakfast, "Paprika," 0:00–0:49.

Through this discussion, students learn to focus their listening toward different elements of a track, familiarize themselves with the instrumental and textural components of a track, and analyze the ways in which different textural components dovetail with one another. A follow-up class activity on "My Baby Just Cares for Me" could rely on the same isolated tracks for the study of form. Students could, for instance, import the instrumental tracks into an audio editing software or DAW and cut and splice the tracks to create a different kind of build-up to accompany the voice. In addition to building familiarity with basic audio editing tasks, this type of activity is a hands-on method to learn about the way form is often delineated through textural changes.

This article explores how source separation can facilitate the analysis of recordings in music theory classrooms focusing on popular music. Music source separation is a technology that converts a recorded song into multiple files, each of which contains different components of the original track. Source separation models extract four discrete components from a recording: vocals, percussion, bass, and "other" (Example 2).[2] The technology enables students to conduct analytical tasks—transcription, harmonic analysis, timbral analysis, or recomposition—without relying on staff notation. Rosa Abrahams writes that upholding the ability to read staff notation as a pre-requisite for enrolling in music theory courses acts as a barrier to entry to several students who may be interested in the topic (2021). In response to Abrahams' call for expanding notions of musical literacy, I frame source separation as a form of musical representation that allows analysts to parse a recording into its subcomponents without significantly losing timbral, pitch, and rhythmic features that may get compromised in a transcription. It can thus be a helpful tool in a variety of contexts in which an instructor may choose to decenter staff notation in favor of recordings: designing a course for non-music majors, teaching popular music genres, and changing the pre-requisites for a core music theory class.

2 For this reason, pop, hip-hop, rock, country, folk, and jazz are all amenable genres to this separation into four tracks. Source separation is less efficient in other contexts. Most instruments in a symphony orchestra would be relegated to the "other" track, for instance, and the entirety of an a cappella choral ensemble would be categorized as "vocals."



Example 2
Flowchart depicting a typical music source separation model.

By introducing source separation in the classroom, moreover, instructors can invite students to critically engage with new technologies. As tools for processing and generating music become increasingly accessible and effective, it is important for students to familiarize themselves with their potential and limitations. Writing about her use of the social media platform TikTok in a music theory assignment, Alyssa Barna writes that "[o]ur current students will never live in a world without these forms of technology. . . .Thus, it is relevant to their work, careers, and daily lives to critically observe how musicians utilize, create, and in some cases, make a living using these platforms, and how the work we do in the theory classroom can relate in small ways" (2024, 22). Following Barna, I suggest that we approach audio processing tools such as source separation in a similar way by fostering a critical understanding of how these tools can impact music analysis and creation. Rather than framing source separation and other artificial-intelligence-based tools as technologies that will fundamentally alter music pedagogy, music theory instructors can view them as one of many forms of musical engagement that students will encounter in their musical education.

In the first section of the article, I outline some situations in which source separation can be productively integrated in music theory and analysis courses focused on popular music. The discussion focuses on timbral analysis and melodic/

harmonic analysis. The second section of the article summarizes recent developments in source separation and outlines some considerations for choosing a source separation model. The conclusion explores how source separation can open the music theory classroom to the perspectives of other areas of music research: audio signal processing, music information retrieval, and interventions on music and artificial intelligence. Introducing audio processing tools in music analysis courses, I contend, is an opportunity for illustrating some of the ethical considerations that arise when using deep-learning-based technologies. This article thus seeks to demystify source separation to explore the ways in which it could facilitate music analytical tasks while encouraging productive discussions on the relationship between music theory pedagogy and audio processing.

Classroom Applications

Source separation tools have a range of practical applications. As we will see in section 3, they are often marketed towards DJs and songwriters. Musicians can use them for remixing, sampling, and creating a karaoke track. Within music studies, the technology has been used to enable the analysis of vocals (Manabe 2019, Duguay 2022, Nobile 2022, de Clercq 2024). For music theory instructors, access to high-quality instrumental and vocal tracks can facilitate some standard writing and ear training tasks. An isolated vocal track can be the basis for a melody harmonization exercise. Students could also practice melody writing or sight-singing with the aid of an instrumental track. Isolated bass and drum tracks can be used to discuss microtiming and groove; separated tracks can also help the transcription of challenging instrumental parts. In a popular music analysis course primarily oriented around recordings, lastly, instructors may wish for students to develop basic audio editing skills such as pitch shifting, slowing down or speeding up, fading, and trimming. Source separation could be taught as one such skill to be used in assignments or in-class activities. In this section, I outline two scenarios that demonstrate the use of source separation technology in teaching timbre, melody, and harmony.

Timbral Analysis

Source separation is particularly effective for the study of timbre. When introducing the topic to my undergraduate students, I use spectrograms along with the opposition table designed by Lavengood (2020) (see Example 3). The table allows students to relate the spectral features of a sound, as they appear on a spectrogram, to verbal qualifiers such as bright/dark, harmonic/inharmonic, and pure/noisy. I might

for instance ask them to compare the extent to which a clarinet sounds "hollower" than a voice, using the spectrogram as a starting point for their description. For many students, seeing how some frequency bands are more pronounced than others on the spectrogram can help them internalize the relationship between the sound they perceive and its acoustic properties. Finally, I teach students to use VAMP plug-ins to extract information—such as the spectral centroid, a marker of timbral "brightness"—from an audio file. These tools for audio feature extraction allow for the further study of timbre.[3] This method allows students to make connections between visualization tools such as spectrograms, their perceptual experience of timbre, and the acoustic features of a sound.[4]

– / + OPPOSITION	BASS 1	MARIMBA
Spectral components - sustain		
bright / dark	–	–
pure / noisy	–	–
full / hollow	–	+
rich / sparse	–	+
beatless / beating	–	–
harmonic / inharmonic	–	+
Spectral components - attack		
percussive / legato	–	–
bright / dark	+	+
Pitch components		
low / high	–	Ø
steady / wavering	–	+

Example 3
Lavengood's opposition table for describing the timbral features of a sound.
Adapted from Example 3 (2020).

3 See Peeters et al. (2011) for a discussion of audio features for timbre-related analysis.

4 Analyzing specific audio features also help illustrate Lavengood's observation that "many aspects of music occur on a gradual scale rather than a simple on/off system" (2020). Tracking a sound's spectral centroid, for instance, can highlight how "brightness" and "darkness" evolve over time.

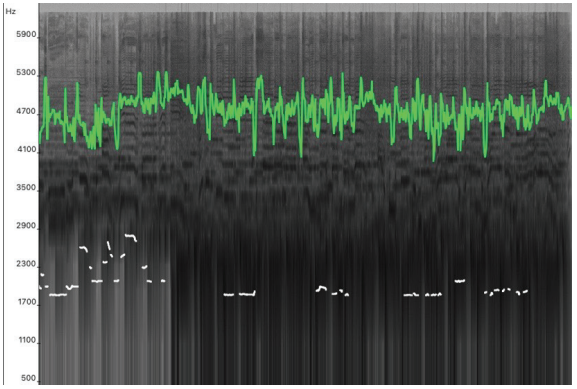
This approach to the study of timbre works well for the analysis of a single sound source. Discussing the timbral features of an instrument couched within a more complicated instrumental texture, however, is more challenging. A student may want, for instance, to analyze the timbral features of Michelle's Zauner's vocal timbre in "Paprika." They may be interested in the way Zauner's timbre becomes brighter and more nasal as her pitch raises in the chorus. Does this piercing effect correlate to a change in the spectral centroid? The harmonicity of her voice? Are certain partials emphasized more than others? To answer these questions, they could supplement their listening with a spectrogram. The image, however, would look too cluttered to make specific observations (Example 4a). Since simultaneous sound sources overlap with one another, it can be difficult to determine which frequencies belong to the voice. The student could also run plug-ins to track the melody, the spectral centroid, and other audio features. This analysis, however, would be applied to all the sound sources—voice, percussion, brass, strings, and synthesizer—in the recording. The melody tracking algorithm, for instance, jumps back and forth between the voice and the bass line.[5] The spectral centroid data, similarly, does not apply to the voice but to the entirety of the track.

Source separation can then be used to help students parse through texturally complex examples. Using an isolated vocal track of Zauner's vocals can provide clearer visual support when analyzing how her timbre evolves throughout the chorus. The isolated track would also allow students to focus their listening exclusively on the voice. The plug-ins would work much more accurately, moreover, because they are now applied only to the isolated vocal track. As shown in Example 4b, the pitch tracking algorithm now accurately tracks the melody, and the spectral centroid data only applies to the vocals. The analytical challenge posed by this example can serve as a springboard for classroom discussions on the limitations of spectrograms and audio feature analysis for analyzing full mixes and encourages students to identify the tools they may need to design an efficient analytical workflow.[6]

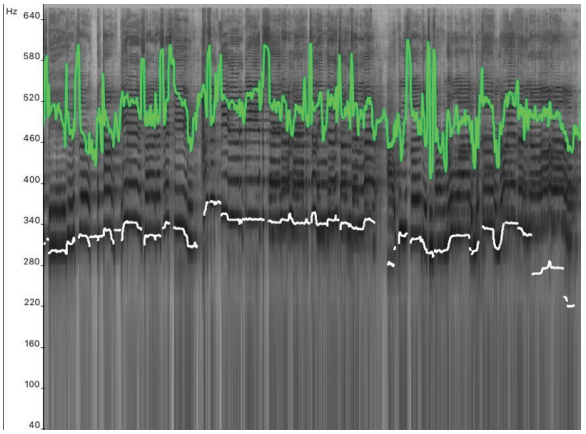
5 I used Mauch & Dixon's PYIN fundamental frequency estimator (2014).

6 Lavengood circumvents this issue in her article by 1) analyzing single pitches and 2) re-recording synthesizer melodies for analysis.

a) Full Mix



b) Isolated vocals



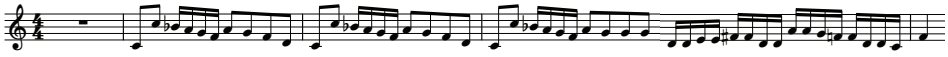
Example 4
Spectrograms for Japanese Breakfast, "Paprika" (2023), 0:25–0:50.

Melodic and Harmonic Analysis

William O'Hara writes that Digital Audio Workstations (DAWs) "shape and reflect the conceptions of musical structure held by the musicians who use them as creative tools" (2025, 84). A DAW's interface encourages a layered approach to musical construction through the vertical stacking of multiple tracks. As a complement to DAW-based pedagogy, source separation can foster a form of layered thinking on which several popular tracks are based. Source separation tools frame vocals, drums, and bass as the main building blocks of recorded music and relegate harmonic instruments to

an "other" category. Harmony is peripheral to this conception of musical organization. Source separation instead encourages a form of contrapuntal thinking that highlights the way different instruments dovetail together to create a full mix.

The Introduction described how access to isolated instrumental layers can facilitate the study of musical texture. The ability to separate a vocal melody from its context in a full track also presents opportunities for the study of melody and harmony in popular music. Students can compose a chord progression to accompany a vocal track, write a complementary melodic line, or study the relationship between melody and harmony. In a music theory course for music majors, for instance, I introduce the isolated vocals for the chorus of Victoria Monét's "Smoke" featuring Lucky Daye (2023). Students then transcribe the melody in staff notation and harmonize it (Example 5). They generally interpret the melody in F major because of the multiple B♭s and the final F.[7] The F♯, which strikes many students as initially disorienting, is then typically harmonized with a secondary dominant or other type of chromatic harmony. Although I could provide students with a transcription of the melody for this exercise, listening to the isolated track allows them to consider Monét's articulation and intonation in their melodic analysis and harmonization.



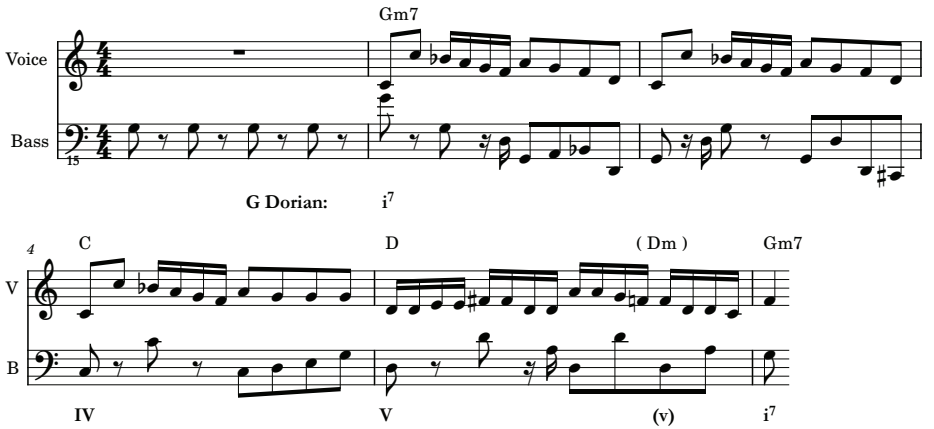
Example 5
Victoria Monét ft. Lucky Daye, "Smoke" (2023). Transcription of vocal melody, 0:13–0:31.

At this point in the lesson, students are very familiar with the vocal track. The transcription in staff notation is associated with a tangible musical performance. When we listen to the full song, this familiarity with the vocals allows them to quickly reinterpret the melody in the context of the Gm^7 – C – D – Dm chord loop (Example 6). The previously surprising F♯ is now contextualized as the third of a major triad. The seemingly mode-defining Cs are non-chord tones or added fourths over Gm^7, and the final F is a chordal seventh. Unlike the harmonizations composed by the students, the song's four-chord progression does not clearly evoke F major. The apparent disconnect between the melody's affordances and the G Dorian mode implied by the groove is an opportunity to discuss some harmonic features of popular music. On the one hand, the song could be interpreted as a textbook example of the melodic-harmonic divorce. It corresponds to Drew Nobile's "hierarchy" divorce, in which "the melody exists at a

7 More rarely, some students opt for a C mixolydian harmonization because of the melodic emphasis on the C.

deeper level of structure than the harmony" (Nobile 2015, 189). In this scenario, the F-major melody expresses the mode of the song while the Dorian-inflected chord loop is a surface-level embellishment, to be interpreted as ii^7 – V –V/ii – v/ii. The F major chord is an "absent tonic," an implied chord that never materializes (Spicer 2017).

On the other hand, we could hear the chord progression as an i^7 – IV – V – v Dorian chord loop. This interpretation positions the G as the song's tonal center. The strong metrical position of Gm^7, the salience of G in the bass line, and the tonic – predominant – dominant organization of the loop work in favor of this hearing. In this scenario, the vocal melody somewhat neutralizes the forward momentum of the chord progression. As a third option, lastly, we could entirely forgo the choice between F major and G Dorian to revel in the tension between the melodic and harmonic layers. This ambiguity confers on the song work a slightly opaque and drifting atmosphere that illustrates the titular smoke. Source separation allows students to experiment with these different interpretations by listening to sound sources in isolation, composing chord progressions and melodies around them, and discussing the way layers work with and against one another to express the mode of the song.



Example 6
Victoria Monét ft. Lucky Daye, "Smoke" (2023). Transcription of vocal melody and bass, 0:13–31.

Lastly, an aural engagement with musical layers can also be used for the development of ear training skills. Students could, for instance, assess the performance of a source separation model. Did the model successfully isolate all aspects (background vocals, reverberated images) of an instrument? Do some unwanted sound sources "bleed" into isolated tracks? Such lines of inquiry could help students think critically about why

a model works or not. A model may not function well, for instance, on other popular genres that do not subscribe as closely to the "vocals, bass, drums, other" paradigm.

A Brief History of Audio Source Separation

The rapid development and increasing accessibility of deep-learning-based tools, including source separation models, for manipulating and generating music raises interesting questions for music theory pedagogy. How, if at all, do these tools change, facilitate, or make obsolete some of music theory instructor's pedagogical goals and approaches? What role, if any, should these tools be given in a music theory curriculum? Music theory instructors can begin to address these questions by critically engaging with these technologies. Such an engagement might entail learning to use these tools, assessing their abilities and limits, understanding the contexts in which they have been created, and weighing the ethical issues surrounding their development and use. To this end, the following section briefly outlines the history, functioning, and types of source separation technologies.[8]

Prior to the development of deep-learning-based technologies for audio source separation, many workarounds were available for extracting isolated vocals or other sound sources from an existing recording. One method was to obtain the stems of an audio file. A "stem" is an audio file containing some components (the vocals, a group of similar instruments, the percussive layer) of a final mix. Stems are eventually combined into a single track to create the final mix. The components of a stem have generally been processed in some way. A vocal stem, for instance, often includes the solo voice and background vocals, processed with compression, EQ, and reverberation. Once a song is commercially released, it becomes nearly impossible to access the original stems because they are encompassed in the final mix's copyright. Since stems are of high interest to DJs and remix artists, however, some platforms like Freesound and ccMixter make available for download free and legal stems licensed under Creative Commons. Legal stems may also be obtained through remix competitions and contests hosted by platforms like Splice and SKIO Music. Although stems may be lacking some aspects of the final mix (such as added reverberation or stereo placement), they were for a long time the most reliable way to access and manipulate the individual components of a released song.[9]

8 For two literature reviews of research on music source separation, see Rafii et al. 2018 and Araki et al. 2025.

9 Because of the scarcity of available stems of commercially recorded popular music, musicians

Audio source separation for music is also an area of research in academia and industry.[10] Throughout the 2000s and 2010s, many of the proposed methods for source separation relied on intrinsic properties of musical recordings. A strand of research on source separation, for instance, relies on the harmonicity of the voice. Some methods isolate the voice by identifying the fundamental frequency of sung pitches and subsequently reconstructing the sound using a series of sine waves. Instead of reconstructing a harmonic sound, other approaches filter out frequencies that do not correspond to the fundamental frequency of the vocal line and its harmonics. The success of these methods is contingent on the voice remaining harmonic through the process, and on the performance of the pitch-identification algorithm. Since pitch-identification algorithms perform better on monodic melodic lines, moreover, vocal harmonies may not be accurately captured in the isolated vocal file. Rather than focusing on the voice, other models rely on the other sound sources in a recording. Repeating Pattern Extraction Technique (REPET) (Rafii & Pardo 2013, 2014) assumes that the musical accompaniment is inherently repetitive. In such cases, the removal of repeating portions of a sound signal would therefore isolate the voice. This approach is especially successful in cases where there are no variations—for instance, a single hi-hat hit in the accompaniment.[11]

---

developed other methods to "unmix" a song. Isolated vocal tracks and instrumental tracks are in particularly high demand given their applicability to karaoke and remixing. Up until the early 2020s, YouTube channels such as MS Project Sound regularly posted isolated vocal tracks of popular songs. Interchangeably named "a cappellas," "pellas," "pells," or "stems," these audio files cater to amateur DJs and producers. Such tracks vary in quality, and were likely obtained through DIY techniques. One method relies on the principle of phase cancellation, in which two identical signals cancel each other out if their waveforms are inverse of one another.

1. Find an instrumental version and a complete version of the song;
2. Align the waveforms of both versions in a digital audio workstation (DAW);
3. Invert the polarity of the instrumental track;
4. Play both tracks at the same time. Only the vocals should be audible.

This approach is quick and somewhat reliable, but it requires access to an instrumental version of the song.

10 The Signal Separation Evaluation Campaign (SiSEC), which ran between 2007 and 2018, was an event that evaluated different methods for source separation submitted by researchers. The publicly available results of each year's campaign ensured continued communication between researchers working on issues of source separation. See Stöter et al. 2018 for the summary of the last SiSEC campaign. The 2021 MDX Challenge (Mitsufuji et al. 2022) and the 2023 Sound Demixing Challenge (Fabbro et al. 2023) were successors to SiSEC.

11 Similar methods include Seetharaman et al. 2017; FitzGerald 2012; Moussallam et al. 2012; Lee & Kim 2015).

Since the late 2010s, the best-performing methods for source separation have relied on deep learning. Deep learning is a form of machine learning that relies on neural networks: computer programs that involve an input layer, multiple hidden layers, and an output layer. Although there are variations between different deep-learning models for source separation, they have all been trained to transform an input (a full mix) into an output (separated tracks). Training a model for source separation requires a dataset containing full audio tracks and their stems. During the training phase, the neural network "learns" from the dataset by adjusting its internal parameters (the hidden layers) to maximize the similarity between its predictions and the original stems. Once a model has been trained in this way, it can make generalizations and separate tracks it has not encountered before.[12]

Upon its release in 2019, Open-Unmix was one of the first high-performing, widely available deep-learning based set of source separation models (Stöter, et al. 2019). Its codes and models are supported by open-source and non-commercial licenses, allowing them to be used for research and personal projects. In the late 2010s, some audio technology companies such as Audionamix[13] and iZotope[14] also released deep-learning based tools for source separation. Beginning in the early 2020s, software and tools for source separation have multiplied and continuously improved in performance. Recent models developed by private-sector companies are often advertised to DJs, producers, and songwriters. Like many new AI-based tools, source separation models can be accompanied by a fair amount of commercial "buzz" or presented as a groundbreaking technological development. Moises AI's website, for instance, tells users that they can now "mute and highlight instruments easily, like [they]'ve always dreamed."[15] Serato, a music production software, promises DJs that they will "drop every jaw in the room

12 See White (2025, 37–44) for an in-depth explanation of neural networks in a musical context. White demonstrates how a neural network may process the melody of "Amazing Grace" as input in order to generate new melodies.

13 Audionamix has not published information on the specific neural network used in XTRAX STEMS, but a conference presentation by Vaneph *et al.* (2016) provides some information on the type of technology developed by Audionamix. A finished mix undergoes various processes—vocal isolation through pitch tracking and separation algorithms based on "state-of-the-art deep learning techniques"—to be broken down in individual sources.

14 iZotope's Music Rebalance tool, introduced in 2018, can also be used to isolate and remix the vocal components of a finished mix. There is no publicly available information on the source separation model or training dataset used by iZotope.

15 See https://Moises.ai/.

when [they] mashup [their] tracks on the fly, with just a few clicks."[16]

Datasets with full mixes and stems are necessary components in the design of deep-learning based source separation models. As mentioned earlier, however, it is difficult to obtain such material without permission from the copyright holders. As a workaround, some researchers have curated publicly available datasets with tracks from a variety of sources. These datasets can be downloaded by anyone and ensure transparency in the research process. MUSDB18, for instance, is a widely used public dataset containing 150 full tracks along with their isolated "drums," "bass," "vocals," and "other" stems (Rafii *et al*. 2017). MUSDB18 is a composite of pre-existing datasets: DSD100 (Liutkus *et al*. 2017), which contains 100 tracks from the "Mixing Secrets" Free Multitrack Download Library, MedleyDB (Bittner et al. 2014), containing tracks from independent artists and recording studios, and miscellaneous tracks licensed under Creative Commons. The dataset was used to train Open-Unmix, and is frequently used in the development of other models. More recently, music AI company Moises released MoisesDB, a dataset of 250 songs and their stems intended to facilitate non-commercial research.

Several source separation models, conversely, are trained with private datasets. Spleeter was trained with an internal dataset from its parent company Deezer, Demucs was trained with a private 3500-song dataset, and Lalal.ai first neural network for source separation was developed with "20 [terabytes] of training data."[17] Such private companies therefore have access to much more data than what is available in MUSDB18 and MoisesDB. The contents these larger datasets, however, is not made public. As Morreale *et al*. (2023) note, this lack of transparency raises questions about the ethics of the dataset. It remains unclear, for instance, whether the creators of songs contained private datasets have been remunerated, acknowledged, or even made aware that their work is being used to train deep-learning models.

Source separation is a relatively new technology, but several authors have already speculated on whether isolated tracks obtained through source separation might eventually be considered copyright infringement (Yoo 2020, Danaher 2022, Coote 2024). For teaching purposes, instructors should ensure that any use of isolated audio files falls under the doctrine of fair use. In my own teaching, I use stems only when they are better suited than the full mix to illustrate a concept, and do not them available for download. When asking students to try a source separation model, I instruct them to use tracks licensed under Creative Commons.

16 See https://serato.com/dj/pro/stems.

17 See https://www.lalal.ai/about/.

Choosing a Source Separation Model

In 2025, musicians have access to dozens of high-performing, source separation models. Some of these tools are free and others require a monthly subscription or one-time purchase. Some models are open source; others have proprietary code that cannot be accessed by the public. Tools can be available on a web browser, through an app, or embedded within a larger platform such as a digital audio workstation. Example 7 compares the features of seven popular models for source separation. This table is not meant to be an exhaustive list of the available tools but is intended to guide the reader wishing to choose a source separation model appropriate for their needs.

When choosing a source separation model, an instructor should weigh performance, cost, usability, ethical considerations. All models in Example 7 perform reliably, and although there may be some variations in the quality of resulting isolated files from one model to the next, they provide highly accurate results. The open-source models are completely free, and the commercial models offer packages ranging from free trials to monthly subscriptions. Usability varies considerably from one model to the next. The open-source models are generally code-based, requiring some experience working with the command line. These models may be desirable for a user wanting to exert more fine-tuned control over the source separation but can also be a barrier to entry. Applications with a basic graphical user interface developed to run source separation models (such as Ultimate Vocal Remover for Demucs) can help circumvent this issue (Example 8). Commercial models, conversely, tend to have clear user interfaces that are accessible to users with no coding experience. Example 9, for instance, shows the Moises user interface. The interface is much more interactive and shares some characteristics of digital audio workstations: the ability to mute and change the volume of tracks, and the ability to see them as they are layered on top of one another. Lastly, instructors should seriously consider the ethical ramifications of using a model trained on proprietary datasets that are not transparent. When using source separation for the classroom, I use Open-Unmix for its transparency and ethical approach to data collection. Demucs (now on its fourth update, Demucs v4) has especially high performance and the ability to be used in a Desktop app. Moises AI's free trial could also be helpful to showcase what a commercial user interface for source separation looks like.

	Cost	Description	Training data	# of stems	Usage
Open-Unmix	Free	- Three pre-trained open-source models: umx, umxhq, and umxl	MUSDB18 (umx) MUSDB18-HQ (umxhq) "private stems dataset" (umxl)	4 (vocals, drums, bass, other)	Command-line/code-based
	Free	- Open-source model	Deezer's private datasets	5 (vocals, piano, drums, bass, other)	Command-line/code-based
Demucs	Free	- Developed at Meta (Facebook) - Open-source model	MUSDB18 and 800-song internal dataset.	4 (vocals, drums, bass, other)	Ultimate Vocal Remover is a free desktop application on which users can run source separation models such as Demucs[1]
Lalal.ai	Free trial (10 min of audio) Lite Pack (90 min, $20) Plus Pack (300 minutes, $27) Pro Pack (500 minutes, $35)	- Proprietary model	Unknown	4 (vocals, drums, bass, other)	Website or app

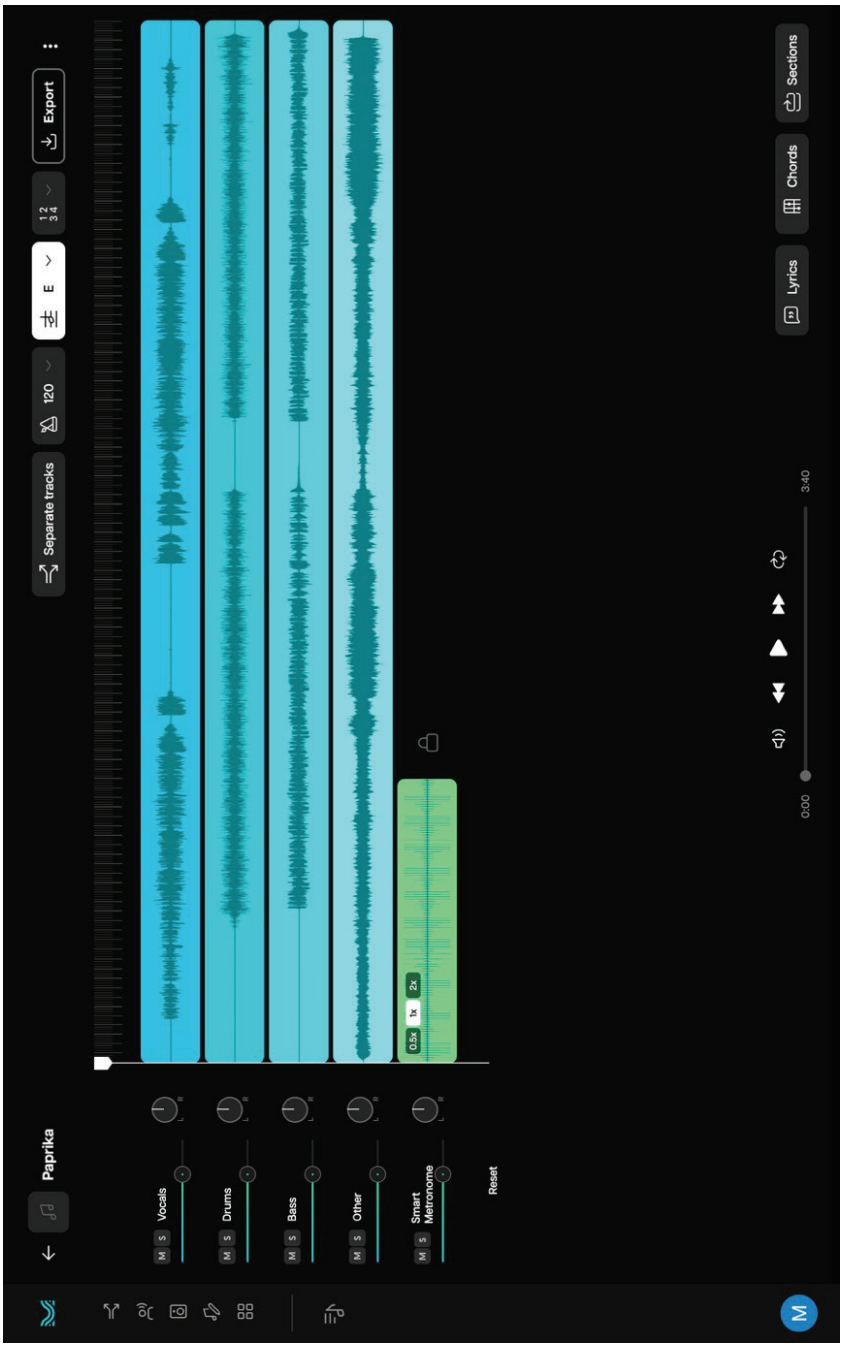
Example 7
Comparison of seven music source separation models.

1 See https://github.com/Anjok07/ultimatevocalremovergui.

	Cost	Description	Training data	# of stems	Usage
Moises	Free version (with restrictions) Moises Premium ($5.99/month) Moises Pro ($29.99/month)	- Available features vary significantly depending on the version. - Uses proprietary model	Unknown	Free version has 4 stems, Pro version has at least 8 (vocals, bass, keyboard, strings, winds, different types of guitars.)	Website or app
Logic Stem Splitter	Available in Logic Pro (free 90 day trial; $199.99 for purchasing software)	- Proprietary model	Unknown	4 (vocals, drums, bass, other)	On Logic Pro 11
Serato Stems	Available in Serato DJ Pro (14-day free trial; $9.99/month or $249 for purchasing software)	- Proprietary model	Unknown	4 (vocals, melody, bass, drums)	On Serato DJ Software

Example 7 (cont'd)
Comparison of seven music source separation models.

Figure 8
Ultimate Vocal Remover user interface for implementing source separation model Demucs v4.



Example 9
Moises AI's user interface for source separation.

Conclusion

This article has outlined some practical uses for integrating audio source separation in the music theory classroom. In addition to facilitating standard ear training and written music theory skills, the technology presents interesting possibilities for the study of timbre and melody. To guide instructors wishing to adopt source separation as a tool, I have provided context on the development of these technologies, as well as ethical issues to consider when choosing a source separation model. To conclude, I discuss how music theory instructors wishing to integrate deep-learning-based tools into classroom discussions and activities could encourage students to develop an understanding of the ways in which these technologies may intersect with their musical lives.

Source separation is a productive entry point into discussions on transparency, data collection, copyright, and accessibility in artificial-intelligence-based tools. Below is a list of potential discussion questions that address different aspects of the technology: its general purpose, transparency and training data, users and accessibility, and performance and bias. These discussion questions could be integrated in class discussions with any of the activities described above. Encountering issues with a finicky user interface, for instance, offers an opportunity to reflect on a tool's accessibility and intended users. For a more formal assignment, students could be assigned a source separation model to research and evaluate as a tool for music analysis.

Topic 1: Introduction to music source separation

- Who built this source separation model?
- What is the purpose of the model, according to its creators?
- Can you think of other uses for this tool?
- In what situations might source separation be useful?
- In what situations might it be harmful, or have unintended consequences?

Topic 2: Transparency and training data

- Can you find out what training data was used to train the model?
- Who created the dataset? Is it available to the public?
- Which artists are included in the dataset? Does it focus on a specific musical genre?
- Did the creators of the dataset seek permission to use this music?

- Do the creators of the model or dataset discuss copyright and intellectual property?
- Given what you've learned about the model's design, would you feel comfortable using it? Why or why not?

Topic 3: Users and accessibility

- Is the model free? Partially free? Available to subscribers only?
- Is the source separation model advertised towards a certain kind of user?
- How might this be reflected in the tool's design?
- Is the user interface intuitive, clear, and accessible? Why or why not?
- Do users need any specialized training to use this model?

Topic 4: Performance and bias

- Was this model designed for a particular musical genre or style?
- Can you find a song on which the model would work well?
- What factors might explain that good performance?
- Can you find a song on which the model would work less well?
- What factors might explain that poorer performance?

Topic 5: Source separation and music analysis

- What kind of music analytical questions can be answered with the aid of source separation?
- Can you think of a situation in which source separation would hinder your analysis?
- In your opinion, what is the place of deep-learning technologies such as source separation tools in a course on music analysis?

The classroom activities and discussion questions listed above are intended to be useful for students with varying levels of familiarity with DAWs, audio editing, or coding experience. When introducing source separation in music analysis courses, I have noticed a range of reactions. Some students are already familiar with the technology and use it to make karaoke tracks, remixes, or simply for listening to components of songs that they like. Others view it as a fun technological trick that can, for instance, be used to evaluate a singer's live vocals. Others are interested in the technology but seem intimidated by the hi-tech trappings of commercial source separation models, especially those that emphasize the use of deep-learning as part

of their marketing strategy. In all these scenarios, source separation can serve as a pedagogical asset. It can facilitate recording-based classroom activities, help students develop musical literacy beyond notation, and teach them to grapple with ethical issues surrounding data transparency and accessibility. If these tools are brought into the classroom with openness and critical inquiry, students can learn to use and question them thoughtfully.

Bibliography

Abrahams, Rosa. 2021. "Rethinking Music Literacy in the Undergraduate Theory Core." Journal of Music Theory Pedagogy 35: 81–108. https://digitalcollections.lipscomb.edu/jmtp/vol35/iss1/2/.

Araki, Shoko, Nobutaka Ito, Reinhold Haeb-Umbach, Gordon Wichern, Zhong-Qiu Wang, and Yuki Mitsufuji. 2025. "30+ Years of Source Separation Research: Achievements and Future Challenges." https://doi.org/10.48550/arxiv.2501.11837.

Barna, Alyssa. 2024. "'Duet Me': Music Theory Pedagogy in the Age of Social Media." Journal of Music Theory Pedagogy 38: 21–44. https://digitalcollections.lipscomb.edu/jmtp/vol38/iss1/3/.

Bittner, Rachel, Justin Salamon, Mike Tierney, Matthias Mauch, Chris Cannam, and Juan Bello. 2014. "MedleyDB: A Multitrack Dataset for Annotation-Intensive MIR Research." In Proceedings of the 15th International Society for Music Information Retrieval (ISMIR) Conference, 155–60. https://doi.org/10.5281/zenodo.1417889.

Coote, Jonathan. 2024. "'Stem-separating' AI is Revolutionising the Music Industry—But at What Cost?" World Intellectual Property Review (WPIR), March 26. https://www.worldipreview.com/future-of-ip/stem-separating-ai-is-revolutionising-the-music-industry-but-at-what-cost.

Danaher, Matthew. 2022. "AI Stem Extraction: A Creative Tool or Facilitator of Mass Infringement?" JTIP Blog, May 3. https://jtip.law.northwestern.edu/2022/05/03/ai-stem-extraction-a-creative-tool-or-facilitator-of-mass-infringement/.

de Clercq, Trevor. 2024. "Some Proposed Enhancements to the Operationalization of Prominence: Commentary on Michèle Duguay's 'Analyzing Vocal Placement in Recorded Virtual Space'." Music Theory Online 30, no. 1. https://mtosmt.org/issues/mto.24.30.1/mto.24.30.1.declercq.html.

Devaney, Johanna. 2022. "Digital Audio Processing Tools for Music Corpus Studies." In The Oxford Handbook of Music and Corpus Studies, edited by Daniel Shanahan, John Ashley Burgoyne, and Ian Quinn. Oxford Academic. https://doi.org/10.1093/oxfordhb/9780190945442.013.11.

Duguay, Michèle. 2022. "Analyzing Vocal Placement in Recorded Virtual Space." Music Theory Online 28, no 4. https://mtosmt.org/issues/mto.22.28.4/mto.22.28.4.duguay.html.

Fabbro, Giorgio, Stefan Uhlich, Chieh-Hsin Lai, Woosung Choi, Marco Martínez-Ramírez, Weihsiang Liao, Igor Gadelha, et al. 2024. "The Sound Demixing Challenge 2023 – Music Demixing Track." Transactions of the International Society for Music Information Retrieval 7, no. 1: 63–84. https://doi.org/10.5334/tismir.171.

FitzGerald, Derry. 2012. "Vocal Separation Using Nearest Neighbours and Median Filtering." In 23rd IET Irish Signals and Systems Conference. https://doi.org/10.1049/ic.2012.0225.

Heidemann, Kate. 2016. "A System for Describing Vocal Timbre in Popular Song." Music Theory Online 22, no. 1. https://www.mtosmt.org/issues/mto.16.22.1/mto.16.22.1.heidemann.html.

Hennequin, Romain, Anis Khlif, Felix Voituret, and Manuel Moussallam. 2020. "Spleeter: A Fast and Efficient Music Source Separation Tool with Pre-Trained Models." Journal of Open Source Software 5 (50): 2154. https://doi.org/10.21105/joss.02154.

Japanese Breakfast. 2021. "Paprika." Track 1 on Jubilee. DOC225, CD.

Lavengood, Megan L. 2020. "The Cultural Significance of Timbre Analysis: A Case Study in 1980s Pop Music, Texture, and Narrative." Music Theory Online 26, no. 3. https://mtosmt.org/issues/

mto.20.26.3/mto.20.26.3.lavengood.html.

Lee, Jun-Yong, and Hyoung-Gook Kim. 2005. "Music and Voice Separation Using Logspectral Amplitude Estimator Based on Kernel Spectrogram Models Backfitting." Journal of the Acoustical Society of Korea 34, no. 3: 227–33. https://doi.org/10.7776/ASK.2015.34.3.227.

Malawey, Victoria. 2020. A Blaze of Light in Every Word: Analyzing the Popular Singing Voice. Oxford University Press.

Manabe, Noriko. 2019. "We Gon' Be Alright? The Ambiguities of Kendrick Lamar's Protest Anthem." Music Theory Online 25, no. 1. https://mtosmt.org/issues/mto.19.25.1/mto.19.25.1.manabe.html.

Mauch, Matthias, and Simon Dixon. 2014. "PYIN: A Fundamental Frequency Estimator Using Probabilistic Threshold Distributions." In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 659–63. https://doi.org/10.1109/ICASSP.2014.6853678.

Mitsufuji, Yuki, Giorgio Fabbro, Stefan Uhlich, Fabian-Robert Stöter, Alexandre Défossez, Minseok Kim, Woosung Choi, Chin-Yun Yu, and Kin-Wai Cheuk. 2022. "Music Demixing Challenge 2021." Frontiers in Signal Processing 1. https://doi.org/10.3389/frsip.2021.808395.

Monét, Victoria. 2023. "Smoke." Featuring Lucky Daye. Track 1 on Jaguar II. RCA–19658 84907 2, CD.

Moore, Allan F. 2012. Song Means: Analysing and Interpreting Recorded Popular Song. Ashgate.

Morreale, Fabio, Megha Sharma, and I-Chieh Wei, "Data Collection in Music Generation Training Sets: A Critical Analysis." 2023. In Proceedings of the 24th Int. Society for Music Information Retrieval (ISMIR) Conference. https://hdl.handle.net/2292/65322.

Moussallam, Manuel, Gaël Richard, and Laurent Daudet. 2012. "Audio Source Separation Informed by Redundancy With Greedy Multiscale Decompositions." In Proceedings of the 20th European Signal Processing Conference (EUSIPCO 2012), 2644–48.

Nobile, Drew F. 2015. "Counterpoint in Rock Music: Unpacking the 'Melodic-Harmonic Divorce'." Music Theory Spectrum 37, no. 2: 189–203.

———. 2022. "Alanis Morissette's Voices." Music Theory Online 28, no. 4. https://mtosmt.org/issues/mto.22.28.4/mto.22.28.4.nobile.html.

O'Hara, William. 2025. "'Let's Think in Layers': On Twenty-First-Century Instruments of Public Music Theory." Music Theory Spectrum 47, no. 1: 83–90.

Peeters, Geoffroy, Bruno L. Giordano, Patrick Susini, Nicolas Misdariis, and Stephen McAdams. 2011. "The Timbre Toolbox: Extracting Audio Descriptors from Musical Signals." The Journal of The Acoustical Society of America 130, no. 5: 2902–16. https://doi.org/10.1121/1.3642604.

Pierard, Tom, and David Lines. 2022. "A Constructivist Approach to Music Education with DAWs." Teachers and Curriculum 22, no. 2: 135–45. https://doi.org/10.15663/tandc.v22i2.406.

Rafii, Zafar, Antoine Liutkus, and Bryan Pardo. 2014. "REPET for Background/Foreground Separation in Audio." In Blind Source Separation: Advances in Theory, Algorithms, and Applications, edited by Ganesh R. Naik and Wenwu Wang, 395–411. Springer.

Rafii, Zafar, Antoine Liutkus, Fabian-Robert Stöter, Stylianos Ioannis Mimilakis, and Rachel Bittner. 2017. "The MUSDB18 Corpus for Music Separation." https://hal.inria.fr/hal-02190845/document.

Rafii, Zafar, Antoine Liutkus, Fabian Robert Stöter, Stylianos Ioannis Mimilakis, Derry FitzGerald, and Bryan Pardo. 2018. "An Overview of Lead and Accompaniment Separation in Music." IEEE/ACM Transactions on Audio Speech and Language Processing 28, no. 8. 1307–35.

https://doi.org/10.48550/arXiv.1804.08300.

Rafii, Zafar, and Bryan Pardo. 2013. "REpeating Pattern Extraction Technique (REPET): A Simple Method for Music/Voice Separation." IEEE Transactions on Audio, Speech and Language Processing 21, no. 1: 73–84. http://doi.org/10.1109/TASL.2012.2213249.

Rouard, Simon, Francisco Massa, and Alexandre Defossez. 2023. "Hybrid Transformers for Music Source Separation." In ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1–5. https://doi.org/10.1109/ICASSP49357.2023.10096956.

Seetharaman, Prem, Fatemeh Pishdadian, and Bryan Pardo. 2017. "Music/Voice Separation Using the 2D Fourier Transform." In 2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. https://doi.org/10.1109/WASPAA.2017.8169990.

Simone, Nina. 1959. "My Baby Just Cares for Me." Track 6 on Little Girl Blue. Recorded December 1957. Bethlehem BS-6028, 1959, Vinyl LP.

Spicer, Mark. 2017. "Fragile, Emergent, and Absent Tonics in Pop and Rock Songs." Music Theory Online 23, no. 2. https://mtosmt.org/issues/mto.17.23.2/mto.17.23.2.spicer.html#nobile_2015.

Stöter, Fabian-Robert, Antoine Liutkus, and Nobutaka Ito. 2018. "The 2018 Signal Separation Evaluation Campaign." In LVA/ICA: Latent Variable Analysis and Signal Separation, 293–305. https://doi.org/10.48550/arXiv.1804.06267.

Stöter, Fabian-Robert, Stefan Uhlich, Antoine Liutkus, and Yuki Mitsufuji. 2019. "Open-Unmix - A Reference Implementation for Music Source Separation." Journal of Open Source Software 4, no. 41. https://joss.theoj.org/papers/10.21105/joss.01667.

Temperley, David. 2007. "The Melodic-Harmonic 'Divorce' in Rock." Popular Music 26, no. 2: 23–42. https://doi.org/10.1017/S0261143007001249.

Vaneph, Alexandre, Ellie Mcneil, François Rigaud, and Rick Silva. 2016. "An Automated Source Separation Technology and Its Practical Applications." In Proceedings of the AES 140th International Convention. http://www.aes.org/e-lib/browse.cfm?elib=18182.

Walzer, Daniel. 2020. "Blurred Lines: Practical and Theoretical Implications of a DAW-Based Pedagogy." Journal of Music, Technology and Education 13, no. 1: 79–94. https://doi.org/10.1386/jmte_00017_1.

White, Christopher W. 2025. The AI Music Problem. Routledge.

Yoo, Noah. 2020. "A Flashy New AI Tool Could Be a Producer's Dream and a Copyright Nightmare." Pitchfork, March 5. https://pitchfork.com/thepitch/a-flashy-new-ai-tool-could-be-a-producers-dream-and-a-copyright-nightmare/.