



Vol. 3 N.º 9 / Guayaquil
mayo 2024
ISSN 2953-657x

Optimizando la mezcla: Answer Set Programming en la producción musical

Flavio Everardo

Escuela de Humanidades y Educación,
Tecnológico de Monterrey Campus Puebla
Puebla, México
Instituto de Ciencias de la Computación,
Universidad de Potsdam
Potsdam, Alemania
flavio.everardo@tec.mx

Ileana Benítez

Escuela de Humanidades y Educación,
Tecnológico de Monterrey Campus Puebla
Puebla, México

Yamil Burguete

Escuela de Humanidades y Educación,
Tecnológico de Monterrey Campus Puebla
Puebla, México

Resumen

En la producción musical, el proceso de mezcla es una tarea compleja donde se deben tomar decisiones interconectadas, tanto técnicas como creativas, para lograr un resultado apropiado. La producción musical inteligente (*Intelligent Music Production*) es un campo emergente que combina métodos cognitivos con procesos de creación y producción musical, que van desde la composición hasta la masterización. Answer Set Programming (ASP) es un paradigma declarativo de programación del área de representación de conocimiento y razonamiento capaz de resolver problemas complejos de optimización, el cual ha demostrado prometedores avances en procesos de producción musical. Este artículo propone un sistema basado en ASP llamado *frmixerr* como un enfoque inteligente para optimizar el proceso de mezcla, con generación automática de mezclas considerando criterios de optimización derivados de un análisis estadístico y una extracción de características en la respuesta en frecuencia. Se realizó una comparación sobre la calidad subjetiva de mezclas generadas por profesionales y por *frmixerr*. No hay diferencias sustanciales en la evaluación subjetiva entre la mezcla generada por el algoritmo y las realizadas por profesionales con diferentes años de experiencia en mezcla.

Palabras clave: Answer Set Programming, *frmixerr*, inteligencia artificial, mezcla automática, producción musical inteligente,

Abstract

In music production, the mixing process is a very complex task where interconnected technical and creative decisions must be made to achieve an appropriate result. Intelligent Music Production is an emergent field that combines cognitive methods with music creation and production processes, ranging from composition to mastering. Answer Set Programming (ASP) is a declarative programming paradigm in the Knowledge Representation and Reasoning area capable of solving complex optimization problems, which has shown promising advances in music production processes. This article proposes an ASP-based system called *frmixerr* as an intelligent approach to optimize the mixing process, automatically generating mixes that consider optimization criteria derived from statistical analysis and frequency response feature extraction. Subjective quality was compared between mixes generated by professionals and *frmixerr*. There are no substantial differences in the subjective evaluation between the mix generated by the algorithm and the ones generated by professionals with different years of mixing experience.

Keywords: Answer Set Programming, *frmixerr*, artificial intelligence, automatic mixing, intelligent music production

1. Introducción

En el contexto de la producción musical, es bien sabido que la mezcla es una tarea muy desafiante y de optimización. Para que una mezcla suene bajo los estándares de la industria, debemos asumir que los pasos previos de producción, como la grabación y su edición correspondiente, se realizaron correctamente para que el ingeniero de mezcla pueda tomar decisiones que implican desafíos técnicos.¹ Algunos de estos retos involucran asegurar el balance entre las fuentes, su audibilidad mediante la minimización del enmascaramiento, la distribución de elementos en el campo panorámico, limitar el rango dinámico o incluso garantizar una amplitud estéreo deseada. Además, existen otros aspectos no técnicos a considerar, como decisiones creativas o artísticas, que derivan decisiones como la aplicación de efectos de audio como reverberación artificial o la automatización de la frecuencia de corte en un filtro.

Las decisiones de mezcla no suelen tomarse de forma aislada sin afectar a otras fuentes o comprometer el resultado. La interacción de estas decisiones de mezcla permite un constante equilibrio en el intento de lograr el mejor resultado posible.² Hoy en día, con la democratización de la tecnología, los productores tienen un mayor acceso a herramientas de producción musical. Ya sea analógico o digital, el encargado de la mezcla realiza los mismos procesos, como ecualización, compresión, distribución panorámica o ajuste del nivel general para optimizar la mezcla, por mencionar algunos. De manera análoga, con el avance de diferentes áreas de la inteligencia artificial (IA) —particularmente enfocándonos en representación del conocimiento y razonamiento (*Knowledge Representation*

and Reasoning; KRR)³; el aprendizaje automático, también conocido como aprendizaje de máquina (*Machine Learning*; ML)⁴ o el procesamiento del lenguaje natural (*Natural Language Processing*; NLP)⁵— contamos con herramientas inteligentes que apoyan las actividades diarias humanas, como los automóviles de conducción o manejo autónomos, asistentes de voz inteligentes e incluso los sistemas de recomendación de música.⁶ En la música y en sus procesos internos, como la composición o la producción, nos hemos encontrado con inclusiones de la IA en todo el proceso de creación musical⁷ desde la composición hasta la masterización. En otras palabras, existen aplicaciones inteligentes en el mundo del audio y de la música debido al concepto de cognificación, el cual busca la aplicación de algoritmos de IA y que ha permitido el desarrollo de campos emergentes como la producción musical inteligente (*Intelligent Music Production*; IMP)⁸.

Para el alcance de este trabajo, presentamos *frmixerr*, un sistema con base en los avances y la injerencia que ha tenido el

3 M. Gelfond y V. Lifschitz, «The Stable Model Semantics for Logic Programming», *ICLP/SLP*, vol. 88 (agosto 1988): 1070–1080.

4 C. M. Bishop y N. M. Nasrabadi, *Pattern Recognition and Machine Learning* (Nueva York: Springer, 2006).

5 Manning, C. y H. Schütze, *Foundations of Statistical Natural Language Processing* (MIT Press, 1999).

6 M. Coeckelbergh, *AI Ethics* (MIT Press, 2020). Sameer Balaganur, «How Spotify's Algorithm Manages To Find Your Inner Groove», *Analytics India Mag*, 6 de enero de 2020, <https://analytic-indiamag.com/how-spotifys-algorithm-manages-to-find-your-inner-groove/>

7 Carolina Sacristán-Ramírez, Flavio Everardo, Yamil Burguete y Bretch De Man, «AI in Music: Implications and Consequences of Technology Supporting Creativity», en *What AI Can Do*, edición de Manuel Cebral-Loureda, Elvira G. Rincón-Flores y Gildardo Sánchez-Ante, 233–252 (Nueva York: Chapman and Hall/CRC).

8 Bretch De Man, Ryan Stables y Joshua D. Reiss, *Intelligent Music Production* (Nueva York: Routledge, 2019).

David Moffat y Mark B. Sandler, «Approaches in Intelligent Music Production», *Arts*, vol. 8, n.º 125 (2019).

A. Palladini, «Intelligent Audio Machines» (charla magistral en 4th Workshop on Intelligent Music Production WIMP-18, Huddersfield, Reino Unido, septiembre de 2018).

Eduardo Reck Miranda (Ed.), *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity* (Springer Nature, 2021).

1 A. Case, *Mix smart: Professional Techniques for the Home Studio* (Taylor & Francis, 2012).

2 M. Senior, *Mixing Secrets for the Small Studio* (Routledge, 2018).

área de KRR en la producción musical inteligente, particularmente mediante uno de sus paradigmas llamado *Answer Set Programming* (ASP).

ASP⁹ es un formalismo basado en reglas para modelar y resolver problemas combinatorios (de optimización) intensos en conocimiento.¹⁰ Ha aumentado rápidamente el número de aplicaciones tanto en el ámbito académico como en la industria, como la planeación, planificación (*scheduling*), diseño de configuraciones, biología, logística¹¹ e incluso la música.¹² ASP y la música han demostrado aplicaciones principalmente en la composición, pero ha habido poco progreso en áreas de postproducción, particularmente en la mezcla.¹³

Retomando nuestra motivación, al ver la mezcla como un problema de optimización complejo, el cual embona perfectamente en ASP, este artículo describe las siguientes contribuciones:

- Presentar la metodología de *frmixerr*, un sistema inteligente, capaz de generar mezclas de forma automática. Este enfoque inteligente propone balances de niveles mediante la comparación de la magnitud general de frecuencias de la mezcla con una referencia. Para nuestros propósitos, usamos el ruido rosa como referencia.
- Demostrar la fiabilidad de nuestra propuesta inteligente mediante la imple-

mentación de una evaluación de escucha subjetiva comparando nuestra solución con mezclas hechas por personas con variedad en sus años de experiencia.

El resto del artículo se compone de la siguiente manera: la sección II presenta los antecedentes necesarios para el desarrollo de *frmixerr*, la que se detalla en la sección III. Las secciones IV y V explican el diseño e implementación de la prueba subjetiva, así como sus resultados, respectivamente. Finalmente, la sección VI cierra el texto discutiendo los resultados, así como futuras investigaciones.

2. Antecedentes

En esta sección, se describen los conceptos necesarios que sustentan nuestra propuesta. Principalmente, nuestra investigación se fundamenta en IMP y ASP.

2.1. Producción musical inteligente

Producción musical inteligente (*Intelligent Music Production*; IMP)¹⁴ es un área de investigación que incorpora IA en procesos de producción y postproducción como grabación, mezcla o *mastering*. En palabras de David Moffat y Mark B. Sandler¹⁵, la IMP tiene la premisa de cambiar fundamentalmente la manera en la que ingenieros y consumidores interactúan con la música, no solo permitiendo la colaboración entre humanos y sistemas inteligentes, sino también explorando y comprendiendo nuevas dimensiones e ideas dentro del espacio de la producción musical. Alineado con los conceptos de IA establecidos por Russell y

9 Gelfond y Lifschitz, «The Stable Model...». Torsten Schaub y Stefan Woltran, «Answer Set Programming Unleashed!», *KI-Künstliche Intelligenz*, vol. 32, n.º 2 (2018): 105-108.

10 Chitta Baral, *Knowledge Representation, Reasoning and Declarative Problem Solving* (Cambridge University Press, 2003).

11 Esra Erdem, Michael Gelfond y Nicola Leone, «Applications of Answer Set Programming», *AI Magazine*, vol. 37, n.º 3 (2016): 53-68.

Andreas Falkner *et al.*, «Industrial Applications of Answer Set Programming», *KI-Künstliche Intelligenz*, vol. 32, n.º 2 (2018): 165-176.

12 Sacristán-Ramírez *et al.*, «AI in Music...».

13 Flavio Everardo, «Over a Decade of Producing Music with Answer Set Programming. A Survey» (charla en Trends and Applications of Answer Set Programming TAASP Workshop, Klagenfurt, Austria, noviembre 2020).

14 De Man, Stables y Reiss, *Intelligent Music Production...*

Moffat y Sandler, «Approaches in Intelligent...».

Palladini, «Intelligent Audio Machines».

Miranda, *Handbook of Artificial...*

15 Moffat y Sandler, «Approaches in Intelligent...».

Norvig¹⁶, estos sistemas inteligentes deben ser capaces de tomar decisiones automáticamente, pensar y razonar sobre esas decisiones, sacar conclusiones y aprender de la experiencia al acumular conocimiento.

Adicionalmente, un sistema general de IMP sigue una arquitectura como se describe en la figura 1. En principio, un sistema inteligente para la producción musical considera diferentes tipos de datos de entrada que se clasifican en materiales sonoros y no sonoros. Estos últimos corresponden a partituras o textos, mientras que los sonoros son archivos de audio. El sistema toma esta información en conjunto con la interacción que viene por parte del usuario para su respectivo análisis. Esto desencadena decisiones validables por el usuario y, en caso de aprobarlo, se realiza una acción específica antes de que el resultado sea devuelto al usuario.

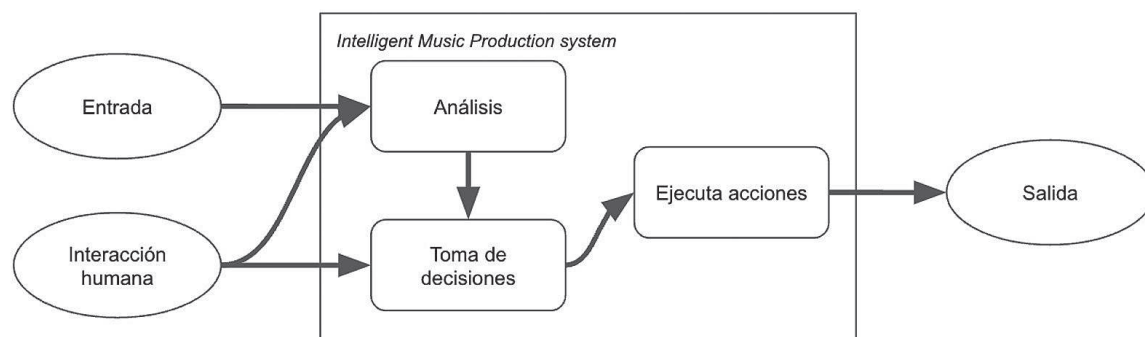


Figura 1: diagrama de un sistema de producción musical inteligente tomado de *AI in Music: Implications and Consequences of Technology Supporting Creativity*¹⁷.

La IA en un sistema de esta naturaleza no se aplica de la misma manera. Se requieren diferentes niveles predefinidos de automatización o grados de participación de la IA para cumplir los propósitos definidos. Un sistema de IMP se puede clasificar en una de estas cuatro áreas¹⁸:

1. **Automático:** un enfoque completamente automatizado donde el sistema

es visto como una caja negra y no se requiere interacción humana. En otras palabras, el sistema toma el control total de los procesos internos y la toma de decisión.

2. **Independiente:** asistente en el que se delegan algunas tareas específicas para proporcionar un resultado cercano al producto terminado, con el usuario actuando como supervisor.
3. **Sugestivo:** la participación del sistema se limita al análisis y la propuesta de recomendaciones, mientras que el usuario tiene control completo sobre el resultado.
4. **Perceptivo:** el sistema proporciona información adicional, datos textuales, visualizaciones, para aconsejar al usuario sobre información o conocimiento sin realizar modificaciones a los datos de entrada.

En otras palabras, la participación de la IA en un sistema de IMP varía desde ser un sistema de asesoramiento que utiliza información y conocimiento, resultando en decisiones más rápidas o mejores para que el usuario las tome, hasta controlar todo el proceso de los datos de entrada siguiendo un enfoque de caja negra. En el contexto de la mezcla, podríamos ejemplificar un sistema de IMP capaz de únicamente recomendar un procesamiento *mid-side* para corregir los problemas de fase en una señal estéreo, hasta el punto de poder realizar una mezcla en su totalidad aplicando balances, paneo, compresión y equalización. En el medio, encontramos niveles que re-

¹⁶ Stuart Russell y Peter Norvig, *Artificial Intelligence: A Modern Approach* (Pearson, 2020).

¹⁷ Sacristán-Ramírez et al., «AI in Music...».

¹⁸ Sacristán-Ramírez et al., «AI in Music...».

lajan el grado de autonomía para resolver problemas específicos, como hacer exclusivamente el balance de las fuentes sonoras o gestionar el rango dinámico de un número de pistas.¹⁹

Diferentes áreas de la IA han sido probadas en diversas tareas dentro del proceso de creación o composición musical. A su vez, la IA ha mostrado avances en el campo de la composición e incluso para la producción musical.²⁰ En este artículo nos centramos en las cualidades que tiene *Answer Set Programming* para integrarse en un sistema de producción musical inteligente.

2.2. Answer Set Programming

ASP es un formalismo basado en reglas para modelar y resolver problemas combinatorios (de optimización) intensos en conocimiento²¹ con orígenes en el área de KRR. Sus raíces parten de la programación lógica, permitiendo que ASP tenga una sólida teoría, así como implementaciones de alto rendimiento. Es por eso que ha demostrado ser exitoso en el ámbito académico como en la industria, resolviendo problemas complejos como planeación, planificación (*scheduling*), diseño de configuraciones, biología, logística²² e incluso la música.²³

ASP tiene más de una década de historia en la música, particularmente en composición,²⁴ con reglas y restricciones específicas para un estilo musical o enfoques independientes del género con conocimiento especializado para cubrir reglas musicales generales. Sin embargo, en tér-

minos de manejo de señales de audio, se ha hecho poco progreso en la postproducción, específicamente en la mezcla. Everardo²⁵ realizó un primer esfuerzo de un sistema inteligente capaz de proponer el balance y paneo utilizando un conjunto de conocimientos sobre las mejores prácticas de mezcla de ingenieros provenientes de la literatura.

Uno de los principales beneficios de ASP como paradigma de lenguaje de programación, en contraste con la programación tradicional, es que nos permite especificar el problema por medio de reglas y restricciones en lugar de codificar la solución de un problema en cuestión. En la práctica, dejamos que un solucionador de ASP (*solver*) busque por medio de inferencias lógicas, aprendizaje y razonamiento, si el problema es satisficible y si tiene una o varias soluciones. En este sentido, podemos enseñar al *solver* a entender las reglas y restricciones para realizar la mezcla en lugar de hacerla por nuestra cuenta.

En la sección siguiente, explicaremos la metodología de cómo nuestro sistema inteligente realiza una mezcla inicial por medio de ASP.

3. *frmixerr*

Esta sección presenta la metodología utilizada para generar mezclas por parte del sistema inteligente llamado *frmixerr*. El sistema, por ahora, se encuentra en la intersección de las categorías de un sistema automático y uno independiente. Se dice que es un sistema automático ya que el usuario no supervisa ni dirige ninguno de los procesos internos y la IA es responsable del análisis, toma de decisiones y la

19 Bretch De Man, «Audio effects 2.0: Rethinking the music production workflow (ADC'17)», video en YouTube, 35:17, noviembre 20 de 2017, <https://www.youtube.com/watch?v=yJcqNR7by9o>.

Moffat y Sandler, «Approaches in Intelligent...».

Palladini, «Intelligent Audio Machines».

Sacristán-Ramírez *et al.*, «AI in Music...».

20 Sacristán-Ramírez *et al.*, «AI in Music...».

21 Baral, *Knowledge Representation...*

22 Erdem, Gelfond y Leone, «Applications of Answer Set Programming».

Falkner *et al.*, «Industrial Applications...».

23 Everardo, «Over a Decade...».

24 Everardo, «Over a Decade...».

25 Flavio Everardo, «Towards an Automated Multi-track Mixing Tool Using Answer Set Programming» (ponencia en 14th Sound and Music Computing Conference, julio de 2017).

realización de las acciones pertinentes. Por otro lado, se puede decir que es un sistema independiente ya que se puede ver como un asistente al que se le ha delegado una actividad específica, por ejemplo, el balance de diferentes pistas de un *multitrack*. Con ello, no se busca que proporcione un resultado final, sino cercano a terminado, con el usuario actuando como supervisor. La finalidad de *frmixerr* a largo plazo es que sirva como un asistente para el proceso de mezcla propiciando un diálogo e intercambio de ideas junto al productor o ingeniero. Es evidente que hoy en día el proceso de la producción musical es hecho en su totalidad por personas en donde el *hardware* o *software* no proporcionan ningún tipo de comunicación o colaboración, así como ningún atributo que explique lo que la IA está realizando.²⁶ Las aplicaciones más apetecibles para los usuarios, en su interés por utilizar IA en la música, son aquellas que, en la práctica, consumen mucho tiempo, son tediosas y de naturaleza técnica: es incierto el efecto que tendrá en el resultado, pero ahorrará tiempo y energía para el músico o productor que quiere centrarse en desarrollos creativos. Estas son las áreas donde la mayoría de los especialistas creativos prefieren la implementación de soluciones de IA. Por otro lado, se oponen mucho a las propuestas no supervisadas de tareas creativas.²⁷ En otras palabras, la mayoría de los profesionales prefieren el uso

de la IA como asistente que mejora su flujo de trabajo para acelerar su proceso creativo.

Tradicionalmente, el ingeniero de mezcla realiza el balance de las diferentes pistas en la sesión ajustando cada uno de los *faders* o atenuadores en una superficie de control, consola de mezcla o dentro de una estación digital de trabajo (*Digital Audio Workstation* o DAW). Sea analógico o digital, la equidad de las fuentes sonoras respeta un balance en las frecuencias de la mezcla de manera similar al comportamiento del ruido rosa (*pink noise*).²⁸ El ruido rosa, también conocido como *1/f*, al momento de ser visto en un analizador con filtros de octava, se puede apreciar que todas las bandas de octava tienen el mismo nivel sonoro y cuyo nivel de presión sonora se caracteriza por una densidad espectral inversamente proporcional a la frecuencia (decreciendo 3 dB por octava) como se puede ver en su periodograma en la figura 2. El ruido rosa es utilizado con diversos fines como en mediciones acústicas, pero también se encuentra en el balance de frecuencias en la música. Voss y Clarke²⁹, demostraron que las fluctuaciones de tono e intensidad en el habla y

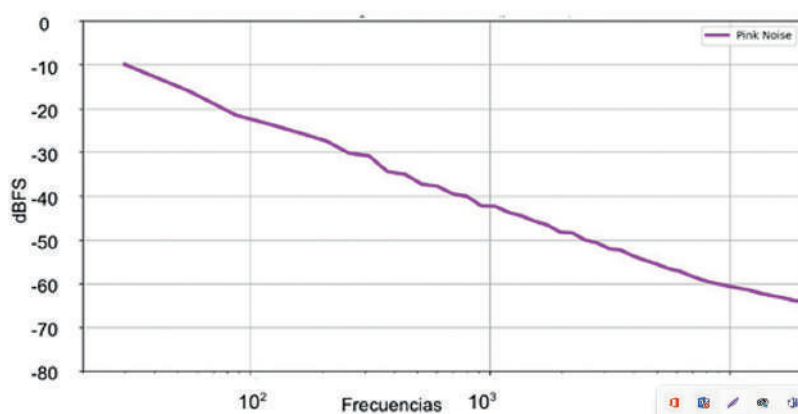


Figura 2: periodograma del ruido rosa. la música corresponden al ruido rosa.

26 David Gunning *et al.*, «XAI—Explainable artificial intelligence», *Science Robotics*, vol. 4, n.º 37 (2019).

Explainable AI (XAI): conjunto de procesos y métodos que permiten a los usuarios humanos comprender y confiar en los resultados creados por IA.

27 Andreas Pfeiffer, *Creativity and technology in the age of AI: Research Report (USA, Europe and Japan)* (Pfeiffer Report, 2018), <https://www.pfeifferreport.com/es-says/creativity-and-technology-in-the-age-of-ai/>

28 Bobby Owsinski, *The Mixing Engineer's Handbook (5th ed.)* (Bobby Owsinski Media Group, 2022).

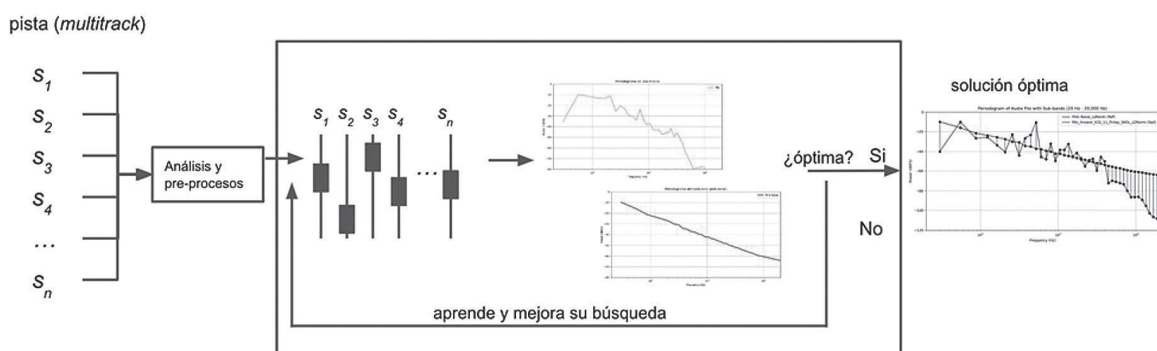
29 Richard F. Voss y John Clarke, « $1/f$ noise" in music and speech», *Nature*, 258 (1975): 317-318.

Richard F. Voss y John Clarke, « $1/f$ noise" in music: Music from $1/f$ noise», *The Journal of the Acoustical Society of America*, vol. 63, n.º 1 (1978): 258-263.

El sistema *frmixerr*, programado en *python* realiza las acciones desplegadas en la figura 3. Primero comienza analizando el *multitrack* dado por el usuario normalizando en pico (*peak normalization*) cada una de sus pistas o señales (s_1, \dots, s_n) para tener la mayor amplitud posible en la mezcla. Después, entra al proceso iterativo en donde el *solver*³⁰ propone un balance en decibeles a escala completa (dBFS) y se realiza la atenuación de cada señal al mismo tiempo de que se genera la suma de señales (Anexo I). Una vez generada la mezcla, se calcula su periodograma y se compara contra el periodograma del ruido rosa. El sistema genera mezclas hasta encontrar la óptima. Se define una solución óptima a la configuración de *faders* cuyo periodograma de dicha mezcla es la más cercana al ruido rosa. Si durante la búsqueda de una mejor solución se encuentra con una diferencia de periodogramas mayor a la que tenga como óptima en

medidas de anchos de filtros utilizadas en el ancho de banda rectangular equivalente (*Equivalent Rectangular Bandwidth*; ERB³¹) el cual es una medida utilizada en el campo de la percepción auditiva para cuantificar la resolución de frecuencia del sistema auditivo humano.

El ERB representa el ancho de banda de un filtro que tiene una forma equivalente al filtro auditivo en una frecuencia central específica. La excitación total que cae dentro de cualquier filtro ERB dado se comprime de manera equivalente, esto resulta en una contribución uniforme al volumen total. Es por esto que ERB proporciona un mecanismo relacionado a la compresión, el enmascaramiento y la sonoridad. Finalmente, nuestro sistema proporciona la configuración en dBFS de cada uno de los *tracks*. La tabla 1 nos muestra los diferentes valores por pista usados en las mezclas mostradas en la figura 4.



ese momento, se descarta esa solución y aprende de esa configuración no fructífera para mejorar su búsqueda. La figura 4 muestra un gráfico de cuatro periodogramas en donde se ilustra la constante mejora acercando la respuesta en frecuencia de la mezcla al ruido rosa.

La diferencia entre periodogramas se hace en 42 bandas siguiendo las me-

Figura 3: arquitectura de *frmixerr*. El sistema genera mezclas buscando refinar en cada interacción la diferencia entre ambos periodogramas (el generado en la mezcla y el ruido rosa) hasta que la solución es óptima. De lo contrario, se descarta y se procede a buscar una mejor opción.

³⁰ Se usó *clingo* como *solver* del grupo de *Potsdam Answer Set Solving Collection* (Potassco). <https://potassco.org/clingo/>

³¹ Brian Glasberg y Brian C. J. Moore, «A Model of Loudness Applicable to Time-Varying Sounds», *Journal of the Audio Engineering Society*, vol. 50, n.º 5 (2002): 331-342.

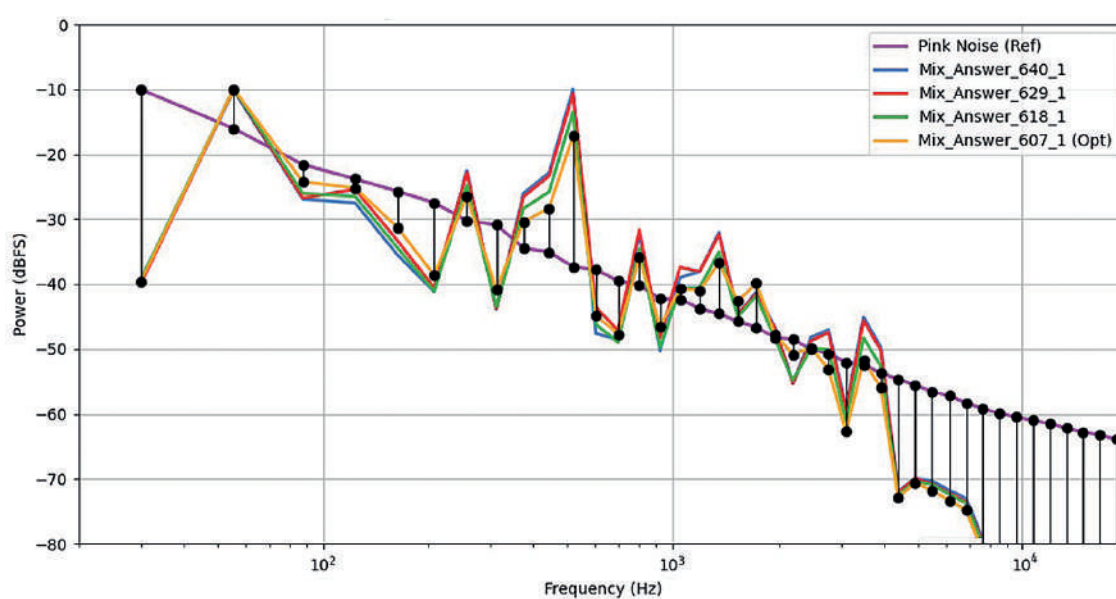
A. J. Simpson, M. J. Terrell y Joshua D. Reiss, «A Practical Step-By-Step Guide to the Time-Varying Loudness Model of Moore, Glasberg, and Baer (1997; 2002)», *Audio Engineering Society Convention 134* (2013).

4. Prueba de escucha-evaluación subjetiva

4.1. Estímulos

Como un primer acercamiento, para tener una referencia inicial del rendimiento de nuestro sistema, se comparó una de las mezclas generadas con mezclas realizadas por una suma simple de señales normalizadas y por cuatro ingenieros de mezcla humanos con diferentes niveles de desarrollo

que no se permitió utilizar ningún efecto o proceso de audio. Para hacerlo más homogéneo, las mezclas resultantes se ajustaron a -23 LUFs para obtener el mismo volumen (*loudness*) y para eliminar el sesgo hacia muestras más fuertes (o más suaves) durante la prueba de audición. Dentro de las múltiples soluciones generadas por *frmixer*, la mezcla a utilizar fue elegida por los autores del presente artículo, previo a tener las mezclas generadas por humanos.



profesional. A cada ingeniero de mezcla se les pidió que hicieran una mezcla de la versión editada de *Same kind of life* por Finlay.³² Dicha mezcla del género pop de 24 segundos de duración consiste en solo el balance en mono de las 16 pistas que contiene el *multitrack*. Esta sesión cuenta con elementos como voz, bajo, guitarras, teclados, bombo, caja y micrófonos de batería (*overheads*). De la misma manera, nuestra mezcla generada por el sistema y la suma de señales también fueron generadas en mono. La razón principal para realizar esta prueba en monoaural, es para evitar distracciones del panorama. De igual forma, cabe resaltar

Figura 4: proceso de optimización de la mezcla. Los valores de 640_1 hasta 607_1 indican la suma de todas las diferencias en dBFS de las 42 bandas. La mezcla Mix_Answer_640_1 (azul) es la más alejada del ruido rosa, mientras las mezclas 629 (roja) y 618 (verde) son resultados más cercanos. La mezcla 4 Mix_Answer_607_1 (amarillo) es la mezcla óptima.

4.2. Diseño y ejecución de la prueba

El método de prueba de escucha que utilizamos fue una prueba de estímulos múltiples con ancla oculta. En otras palabras, se trata de una prueba MUSHRA (estímulo múltiple con ancla y referencia oculta³³) excepto que en este caso no había referencia. El ancla

³² *Multitracks* disponibles bajo una licencia Creative Commons, <https://www.cambridge-mt.com/ms/mtk/>

³³C. Völker *et al.*, «Modifications of the Multi Stimulus Test with Hidden Reference and Anchor (MUSHRA) for Use in Audiology», *International Journal of Audiology*, vol. 57, n.º 3 (2018): S92-S104.

oculta es la suma monofónica del audio sin procesar, donde cada canal ha sido normalizado en pico. La intención de esta mezcla es que sea identificable como la más baja de todas. Sin embargo, es posible que incluso sujetos expertos en el área califiquen esta ancla por encima de la calificación menor o incluso aceptable dependiendo de la percepción de las otras mezclas.³⁴ Para esta prueba se utilizó una escala de nueve valores en total a través de cinco identificadores: mala, pobre, razonable, buena y excelente. Los otros cuatro valores son las ponderaciones intermedias entre cada identificador. La escala se define con valores desde 1 para una mezcla considerada como mala hasta 9 para una excelente.

El orden de las canciones, así como su identificador fueron aleatorios para evitar cualquier tipo de sesgo o diferencia de interpretación de los temas relacionada con el orden de reproducción. El estudio consistió de una sola sesión *offline* con una duración de entre 10 y 20 minutos. Es decir, los audios fueron puestos a disposición de los evaluadores para su escucha y juicio.

Un total de 18 profesionales de la industria (3 mujeres y 15 hombres) con diferentes perfiles, pero con experiencia en mezcla fueron invitados a participar en esta prueba. A cada participante se les pidió llenar un consentimiento informado dando por hecho la realización del ejercicio utilizando audífonos o monitores de estudio. La edad media de los participantes fue de 31.5 años y ninguno identificó sufrir algún padecimiento auditivo. Se pidió a los sujetos su calificación de las seis mezclas del mismo fragmento de canción de acuerdo con lo que para cada quien era sinónimo de calidad. Los sujetos no tenían ninguna información sobre el contenido del audio o el objetivo de la investigación antes de realizar la prueba.

³⁴ Bretch De Man y Joshua D. Reiss, «A Knowledge-Engineered Autonomous Mixing System», *Audio Engineering Society Convention*, 135 (2013).

	Mezclas generadas por <i>frmixerr</i> . (dBFS)			
	640_1	629_1	618_1	607_1
Pistas del <i>multitrack Same kind of life</i> por Finlay				
01_Kick.wav	-28	-28	-28	-28
02_KickSample.wav	-30	-30	-30	-30
03_SnareSample.wav	-27	-27	-27	-27
04_Tom1.wav	-28	-28	-28	-28
05_Tom2.wav	-28	-20	-21	-29
06_Tom3.wav	-30	-30	-30	-30
07_Cymbal.wav	-27	-27	-27	-27
08_Bass.wav	-27	-27	-27	-27
09_ElecGtr1.wav	-30	-30	-30	-30
010_ElecGtr2.wav	-27	-27	-27	-27
011_ElecGtr3.wav	-30	-30	-30	-30
012_Piano1.wav	-28	-28	-28	-28
013_Piano2.wav	-30	-21	-28	-28
014_Organ.wav	-30	-30	-23	-22
015_LeadVox.wav	-18	-18	-21	-22
016_BackingVox.wav	-27	-27	-27	-27

Tabla 1: ejemplo del proceso de optimización de los periodogramas mostrados en la figura 4. El sistema busca reducir la diferencia en dBFS entre su periodograma con respecto al del ruido rosa. Se puede notar entre los cambios más sobresalientes entre las mezclas 607 y 640, que 12 de las 16 pistas no ajustaron sus valores del *fader*. Sin embargo, es notorio el ajuste en las pistas Piano2, Organ y LeadVox.

5. Resultados

Se tomaron las calificaciones para cada una de las seis mezclas por parte de los 18 evaluadores. La mezcla 2 fue generada por nuestro sistema *frmixerr* y, como se mencionó en la sección anterior, usamos un ancla negativa identificada como la mezcla 5. En otras palabras, se puede interpretar como una manipulación experimental con expectativa de resultados bajos. Las mezclas 1, 3, 4 y 6 corresponden a las producciones realizadas por humanos especialistas en mezcla con diversa variedad en sus años de experiencia (desde seis meses hasta diez años). Existen trabajos³⁵ en los que hacen énfasis en que es posible que el ancla pueda ser evaluada al mismo nivel que los otros estímulos.

La escala de sus calificaciones se encuentra entre la calificación mala, con un

³⁵ De Man y Reiss, «A Knowledge-Engineered...».

valor de 1, y excelente, con un valor de 9. El análisis de datos se realizó de la siguiente manera: se calculó la estadística descriptiva pertinente (ver figura 5), posteriormente se realizó un análisis de varianza (ANOVA) de una vía para detectar las diferencias entre grupos. Todo el trabajo de análisis estadístico se realizó con el *software* Jamovi³⁶, que es un programa basado en el lenguaje R³⁷.

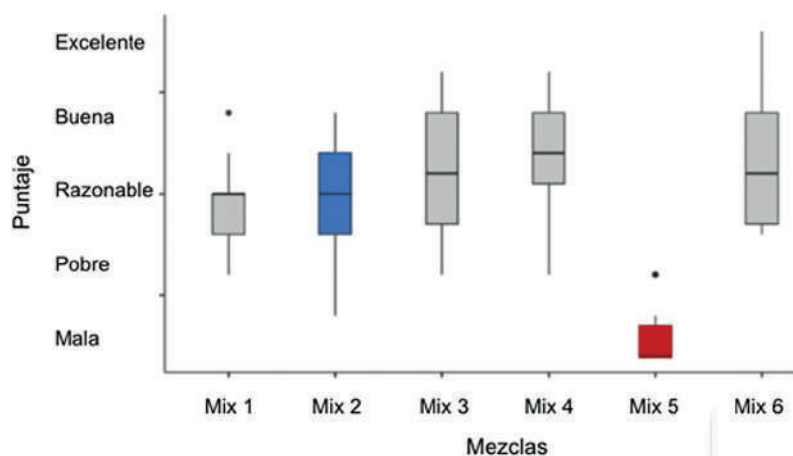


Figura 5: se presentan los diagramas de caja de la evaluación para cada una de las mezclas. La mezcla 5 (rojo) es una manipulación experimental con un resultado bajo esperado (ancla). La mezcla 2 (azul) fue producida por nuestro sistema *fmixer*.

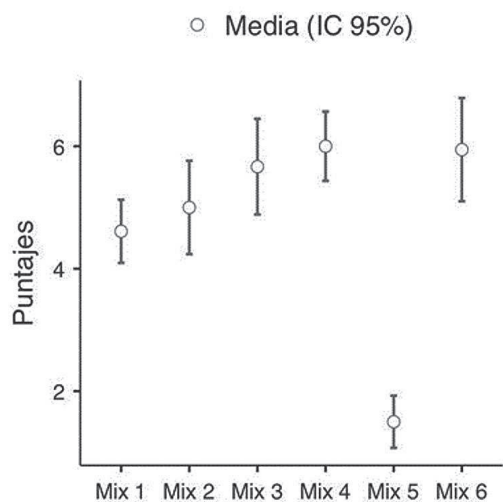


Figura 6: se observan los puntajes medios por mezcla con los intervalos de confianza a 95 %. Se puede observar que la mezcla 5 se encuentra distante del resto de las medias.

Para el análisis descriptivo, se calculó el promedio y desviación estándar de la evaluación de cada mezcla (ver tabla 2). En general se observa que las medias se mantienen entre los valores de 4.6 hasta 6.0 salvo por la mezcla ancla. Este control experimental sirve para ver que todos los evaluadores tienen la capacidad de identificación de las mezclas además de un control de atención. Por su parte, la mezcla de nuestro sistema (mezcla 2), se encuentra aproximadamente al nivel de las otras mezclas producidas por especialistas humanos. De igual

manera, las desviaciones estándares se encuentran en el rango entre 0.85 y 1.69. Con los resultados anteriores se procedió al análisis inferencial para ver si las diferencias detectadas en la variación de las medias son significativas. Por las condiciones de la muestra (18 observaciones por cada mezcla) se decidió correr algunas pruebas para verificar los supuestos de la prueba estadística. Por los valores obtenidos en la prueba de Shapiro-Wilks ($W = 0.981$, $p\text{-valor} = 0.135$), se puede asumir que la muestra presenta un comportamiento normal. En el caso de la prueba de Levene (Fox y Weisberg, 2020) de igualdad de varianzas ($F = 3.87$, $gl_1 = 5$, $gl_2 = 102$, $p\text{-valor} = 0.003$) se identificó que las varianzas son desiguales, por lo tanto, el ANOVA de un factor utilizó la corrección de Welch para desigualdad de varianzas.

36 The Jamovi Project, «jamovi (Versión 2.3) [Computer Software]», Jamovi, 2022, <https://www.jamovi.org>.

37 R Core Team, «R: A Language and environment for statistical computing. (Version 4.1) [Computer software]», Cran.r Project, 2021, <https://cran.r-project.org> («R packages» extraídos de MRAN snapshot 2022-01-01).

Mezcla	Mezcla	Mezcla
Mix1	4.6	1.0369
Mix2	5.0	1.5339
Mix3	5.7	1.5718
Mix4	6.0	1.1376
Mix5	1.5	0.8575
Mix6	5.9	1.6968

Tabla 2. Se observan los valores de las medias y desviaciones estándares para cada una de las mezclas. No hubo datos faltantes, todos los evaluadores calificaron todas las mezclas. La mezcla 2 es la generada por el algoritmo y la mezcla 5 es la mezcla ancla con calificación baja esperada.

Al realizar el ANOVA de un factor, se encontró que existe al menos una diferencia entre las mezclas ($F = 50.2$, $gl1 = 5$, $gl2 = 47.1$, $p\text{-valor} < 0.001$). Por lo anterior, se decidió graficar los valores de las medias con sus intervalos de confianza al 95 % (ver figura 6). Como se puede observar, la mezcla 5 tuvo un comportamiento significativamente diferente al resto de los grupos, todas las comparaciones por pares tuvieron un $p\text{-valor} < 0.001$. La única diferencia significativa entre el resto de las mezclas fue entre la mezcla 1 y la 4 (diferencia = -1.39 , $p\text{-valor} = 0.007$), las cuales correspondieron a las preferencias de los evaluadores a dos mezclas realizadas por ingenieros de mezcla.

Se puede concluir que no existió una diferencia en la evaluación subjetiva entre las mezclas realizadas por profesionales y la mezcla realizada con el algoritmo. Los resultados también muestran que los evaluadores identificaron y calificaron adecuadamente la mezcla 5, que es una manipulación con baja respuesta esperada. Por lo tanto, la mezcla generada por *frmixerr* tiene una aceptación similar a la de profesionales con diferentes años de experiencia en mezcla.

6. Discusión

ASP ha demostrado su viabilidad en la resolución de un problema de optimización de alta complejidad como la mezcla, posicionándose como una alternativa dentro del área de producción musical inteligente. Para el alcance de este trabajo, se propuso un sistema llamado *frmixerr* para generar soluciones balanceadas siguiendo como criterio de optimización una respuesta en frecuencia. Para este caso, el ruido rosa. El sistema es capaz de entregar una configuración inicial en un formato legible para humanos para que cualquier ingeniero de

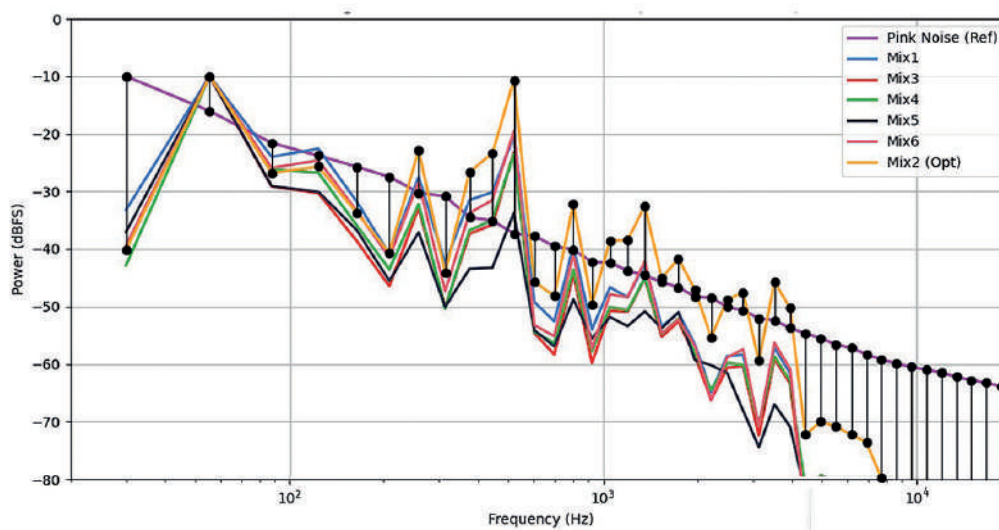


Figura 7: comparación de periodogramas de las seis mezclas utilizadas en la prueba. Cabe mencionar que la mezcla 2 es la más cercana al ruido rosa. La mezcla ancla (5) es la más alejada a la curva de la respuesta objetivo.

audio especializado comience a abordar la mezcla. Los resultados de la evaluación subjetiva demostraron criterios de elección variados, pero coincidiendo en cuatro de seis

mezclas estando la mezcla generada por *frmixerr* dentro de las elegibles y haciendo énfasis en el gusto del evaluador sin ningún otro criterio adicional. En otras palabras, los resultados avalan que nuestra mezcla es igual de probable de poder ser elegida como cualquiera generada por humanos con años de experiencia. No existe una diferencia significativa con respecto a las generadas por humanos.

Es importante destacar que este primer experimento, no da una resolución definitiva, sino que motiva a extender la prueba para conocer el potencial de una herramienta inteligente en la mezcla. Para ello, se listan algunas de las consideraciones inmediatas: ampliar el número de *multitracks*; utilizar un ambiente controlado de escucha; extender *frmixerr* con la posibilidad de generar mezclas que tomen en cuenta otros procesos de mezcla, como paneo, amplitud estéreo, minimización de enmascaramiento, rango dinámico, entre otros. Para ello, es inminente expandir el conjunto de módulos de medición, para permitir una mejor escucha y procesamiento, así como aplicar la correlación cruzada para medir cancelaciones de fase y amplitud estéreo y continuar con el modelo de bandas ERB para medir y minimizar enmascaramiento, por mencionar algunos.

Para este trabajo, se utilizó el ruido rosa como la curva de frecuencia estándar. Sin embargo, vale la pena analizar diferentes respuestas en frecuencia por género. La mezcla generada por *frmixerr* fue la más cercana a la referencia. A pesar de que influyó de forma positiva en el resultado, no

significó que fuese la mejor opción. Combinar otros procesos de audio puede ayudar a moldear la respuesta en frecuencia de la mezcla con una curva referente.

En cuanto al diseño y ejecución de la prueba subjetiva, es necesario brindar a los sujetos con más criterios de evaluación que un simple «me gusta» como respuesta a la subjetividad de la calidad sonora. Debe ser posible poner en juicio otros criterios como el rango dinámico, amplitud estéreo, claridad en las fuentes sonoras, entre otros. De igual manera, nuestro sistema es capaz de generar muchas mezclas válidas. Para garantizar una mejor representación, es necesario realizar una evaluación colegiada previa de mezclas generadas por la IA y no como selección arbitraria de los autores.

Finalmente, se deja para futuras investigaciones evaluar el rendimiento del sistema para demostrar la versatilidad de ASP en la producción musical. Es decir, aclarar dudas sobre cuánto tiempo tarda en generar estas mezclas. Informalmente se puede decir que el sistema genera el primer resultado en cuestión de pocos segundos, aunque la mezcla óptima puede tardar algunos minutos.

7. Agradecimientos

Los autores agradecen a los ingenieros de mezcla Jay Castro, Cesar Arturo Díaz Ordaz, Sebastián Portillo y Guillermo Andrés Vega Castellanos, así como a las y los 18 profesionales de la industria que participaron en la evaluación subjetiva.

Referencias

- Balaganur, Sameer. «How Spotify's Algorithm Manages To Find Your Inner Groove». *Analytics India Mag*, 6 de enero de 2020. <https://analyticsindiamag.com/how-spotifys-algorithm-manages-to-find-your-inner-groove/>
- Baral, Chitta. *Knowledge Representation, Reasoning and Declarative Problem Solving*. Cambridge University Press, 2003.
- Bishop, C. M. y N. M. Nasrabadi. *Pattern Recognition and Machine Learning*. Nueva York: Springer, 2006.
- Case, A. *Mix smart: Professional techniques for the home studio*. Taylor & Francis, 2012.
- Coeckelbergh, M. *AI Ethics*. MIT Press, 2020.
- De Man, Bretch. «Audio effects 2.0: Rethinking the music production workflow (ADC'17)». Video en YouTube, 35:17. Noviembre 20 de 2017. <https://www.youtube.com/watch?v=yJcqNR7by90>
- De Man, Bretch y Joshua D. Reiss. «A Knowledge-Engineered Autonomous Mixing System». *Audio Engineering Society Convention*, 135 (2013).
- De Man, Bretch, Ryan Stables y Joshua D. Reiss. *Intelligent Music Production*. Nueva York: Routledge, 2019.
- Erdem, Esra, Michael Gelfond y Nicola Leone. «Applications of Answer Set Programming». *AI Magazine*, vol. 37, n.º 3 (2016): 53-68.
- Everardo, Flavio. «Towards an Automated Multitrack Mixing Tool Using Answer Set Programming». Ponencia en 14th Sound and Music Computing Conference. Julio de 2017.
- . «Over a Decade of Producing Music with Answer Set Programming. A Survey». Charla en Trends and Applications of Answer Set Programming (TAASP) Workshop, Klagenfurt, Austria, noviembre 2020.
- Falkner, Andreas, Gerhard Friedrich, Konstantin Schekotihin, Richard Taupe y Erich C. Teppan. «Industrial Applications of Answer Set Programming». *KI-Künstliche Intelligenz*, vol. 32, n.º 2 (2018): 165-176.
- Fox, J. y S. Weisberg. «*car: Companion to Applied Regression*. [R package]». 2020. <https://cran.r-project.org/package=car>.
- Gebser, M., R. Kaminski, B. Kaufmann y T. Schaub. «Answer Set Solving in Practice». *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 6, n.º 3 (2012): 1-238.
- Gelfond, M. y V. Lifschitz. «The Stable Model Semantics for Logic Programming». *ICLP/SLP*, vol. 88 (agosto 1988): 1070-1080.
- Glasberg, Brian y Brian C. J. Moore. «A Model of Loudness Applicable to Time-Varying Sounds». *Journal of the Audio Engineering Society*, vol. 50, n.º 5 (2002): 331-342.
- Gunning, David, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf y Guang-Zhong Yang. «XAI—Explainable artificial intelligence». *Science Robotics*, vol. 4, n.º 37 (2019).
- Lewis, M., Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy y L. Zettlemoyer. *Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension*. arXiv preprint arXiv:1910.13461. (2019).
- Manning, C. y H. Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.
- Miranda, Eduardo Reck (Ed.). *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*. Springer Nature, 2021.
- Moffat, David y Mark B. Sandler. «Approaches in Intelligent Music Production». *Arts*, vol. 8, n.º 125 (2019).
- Owsinski, Bobby. *The Mixing Engineer's Handbook (5th ed.)*. Bobby Owsinski Media Group, 2022.
- Palladini, A. «Intelligent Audio Machines». Charla magistral en 4th Workshop on Intelligent Music Production (WIMP-18), Huddersfield, Reino Unido (vol. 14), septiembre de 2018.
- Pfeiffer, Andreas. *Creativity and technology in the age of AI: Research Report (USA, Europe and Japan)*. Pfeiffer Report, 2018. <https://www.pfeifferreport.com/essays/creativity-and-technology-in-the-age-of-ai/>

- R Core Team. «R: A Language and environment for statistical computing. (Version 4.1) [Computer software]». Cran.r Project, 2021. <https://cran.r-project.org>. («R packages» extraídos de MRAN snapshot 2022-01-01).
- Russell, Stuart y Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson, 2020.
- Sacristán-Ramírez, Carolina, Flavio Everardo, Yamil Burguete y Bretch De Man. «AI in Music: Implications and Consequences of Technology Supporting Creativity». En *What AI Can Do*. Edición de Manuel Cebal-Loureda, Elvira G. Rincón-Flores y Gildardo Sánchez-Ante, 233-252. Nueva York: Chapman and Hall/CRC.
- Schaub, Torsten y Stefan Woltran. «Answer Set Programming Unleashed!». *KI-Künstliche Intelligenz*, vol. 32, n.º 2 (2018): 105-108.
- Senior, M. *Mixing secrets for the small studio*. Routledge, 2018.
- Simpson, A. J., M. J. Terrell y Joshua D. Reiss. «A Practical Step-By-Step Guide to the Time-Varying Loudness Model of Moore, Glasberg, and Baer (1997; 2002)». *Audio Engineering Society Convention 134* (2013).
- The Jamovi Project. «jamovi (Versión 2.3) [Computer Software]». Jamovi, 2022. <https://www.jamovi.org>.
- Voss, Richard F. y John Clarke. «“1/f noise” in music and speech». *Nature*, 258 (1975): 317-318.
- . «“1/f noise” in music: Music from 1/f noise». *The Journal of the Acoustical Society of America*, vol. 63, n.º 1 (1978): 258-263.
- Völker, C., T. Bisitz, R. Huber, B. Kollmeier y S. M. Ernst. «Modifications of the Multi Stimulus Test with Hidden Reference and Anchor (MUSHRA) for Use in Audiology». *International Journal of Audiology*, vol. 57, n.º 3 (2018): S92-S104.

ANEXO 1

La programación de programas bajo el paradigma de Answer Set Programming sigue una metodología declarativa. Esto quiere decir, que, en lugar de programar la solución al problema, es necesario establecer las reglas y restricciones dentro del dominio de dicho problema para que el *solver* sea capaz de decirnos si no existe solución alguna, si existe una o diversas. El objetivo de este artículo no es introducir el lenguaje ni sus diferentes constructores, sino ejemplificar la factibilidad de modelar un problema de manera declarativa en una forma compacta y concisa. Para mayor detalles sobre cómo modelar un programa, así como los fundamentos de ASP, puede revisar Gebser *et al.*³⁸

A continuación, se enlista el programa utilizado para proponer las mezclas usando la sintaxis del sistema *clingo*.

Las primeras dos líneas indicarán el número de pistas y si existe preferencia por alguna de estas pistas en particular como el *track* principal o *lead*. Las líneas 4 y 6 representan el conocimiento del número total de pistas en el *multitrack*, así como el dominio de decibeles (dBFS). La línea 8 es la responsable de asignar para cada track T, un valor DB en decibeles y asignarlo a *track_decibel*. En este momento, cada pista ya cuenta con un valor definido de decibeles para escalar su amplitud. La línea 10 es una restricción que indica que no puede haber una pista con una amplitud mayor que la denominada *lead_track*.

La línea 12 por medio de una función externa, convierte cada valor de decibel en amplitud lineal. Finalmente, se busca que la mezcla cuente con los valores de amplitud más elevados posibles, es por ello que la amplitud lineal de cada pista deberá estar entre -3 y 0 dBFS representados por 0.7 y 1.0 en amplitud lineal. Dado las restricciones del lenguaje, se manejan números enteros.

³⁸ M. Gebser et al., «Answer Set Solving in Practice», *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 6, n.º 3 (2012): 1-238.

```
1 |#const tracks = 2.
2 |#const lead_track_id = 1.

4 |track(1..tracks).

6 |decibel(-40..0).

8 |1 { track_decibel(T,DB) :
decibel(DB) } 1
   :- track(T).

10|:- track_decibel(T,DB1),
T!= lead_track_id,

track_decibel(lead_track_id,DB
2), DB1>DB2.

12|track_amplitude(T,
@db2linear(DB)) :-
   track_decibel(T,DB).

14|:- 100 < #sum{A,T :
track_amplitude(T,A)}.
15|:- 70 > #sum{A,T :
track_amplitude(T,A)}.
```