

**PONTIFICIA UNIVERSIDAD
CATÓLICA DEL PERÚ**

FACULTAD DE ARTES ESCÉNICAS



Vocaloid como herramienta para la composición de canciones
y su impacto en el público de la ciudad de Lima entre los 18 y
48 años de edad

Tesis para obtener el título de Licenciado en Música que presenta:

Leandro Gabriel Torres Guerrero

Asesor:

Cesar Humberto Vega Zavala


Lima, 2025

Informe de Similitud

Yo, **Cesar Humberto Vega Zavala**, docente de la Facultad de Artes Escénicas de la Pontificia Universidad Católica del Perú, asesor de la tesis de investigación titulada *Vocaloid como herramienta para la composición de canciones y su impacto en el público de la ciudad de Lima entre los 18 y 48 años de edad*, del autor **Leandro Gabriel Torres Guerrero**, dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de **9%**. Así lo consigna el reporte de similitud emitido por el software *Turnitin* el **04-nov-2024**.
- He revisado con detalle dicho reporte y la tesis, y no se advierte indicios de plagio.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lugar y fecha: Lima, 02 de febrero de 2026

Nombres y apellidos del asesor: Cesar Humberto Vega Zavala	
DNI: 40340759	Firma: 
ORCID: https://orcid.org/0000-0003-2283-5773	

Resumen

Al hablar de “instrumentos virtuales”, no se consideran a los “instrumentos de canto virtual” con la misma jerarquía ya que evidencian la falta de similitud en comparación con un cantante real si se somete al darle el mismo uso para replicar las características propias del canto humano. La capacidad de poder imitar el canto humano a través de un ordenador ofrece una nueva herramienta para la composición musical. Existen *softwares* que se encargan de emular el canto humano como lo es VOCALOID, sintetizador vocal creado por Yamaha Corporation. El objetivo de esta investigación consiste en comprender los elementos del *software* VOCALOID que se consolidan como una herramienta para la composición de canciones como medio de expresión musical y en analizar en cómo la ejecución del mismo impacta en los oyentes. Con esta finalidad, la pregunta de investigación es la siguiente: ¿de qué manera el uso del sintetizador VOCALOID se consolida como una herramienta para la composición de canciones y cuál es su impacto en el público de la ciudad de Lima entre los 18 a 48 años de edad? Para hablar de cómo VOCALOID aporta a la composición, se analiza cómo se ha utilizado el *software* en dos canciones seleccionadas. Para medir el impacto, realizaron encuestas a distintas personas que se encuentren dentro del rango desde los 18 años hasta los 48 años de edad. Las respuestas tratan temas sobre la importancia del realismo de la voz virtual, el contenido audiovisual (*videoclip*) y la preferencia de los encuestados por una canción en la cual la voz virtual tiene una sonoridad más artificial. La definición de lo que es considerado “real” está enlazado con el contexto musical en el cual se utilizan los parámetros de los cantantes virtuales para complementarse con el estilo musical.

Palabras clave: composición, vocaloid, síntesis, síntesis vocal, música, vst

Abstract

When talking about “virtual instruments”, “virtual singing instruments” are not considered with the same hierarchy since they show the lack of similarity compared to a real singer if it is subjected to giving it the same use to replicate its own characteristics. of human song. The ability to imitate human singing through a computer offers a new tool for musical composition. There are softwares that are responsible for emulating human singing, such as VOCALOID, a vocal synthesizer created by Yamaha Corporation. The objective of this research is to understand the elements of the VOCALOID software that are consolidated as a tool for song composition as a means of musical expression and to analyze how experimentation with vocal synthesis impacts listeners. With this purpose, the research question is the following: in what way is the use of the VOCALOID synthesizer consolidated as a tool for song composition and what is its impact on the public in the city of Lima between 18 and 48 years old? old? To talk about how VOCALOID contributes to the composition, we analyze how the software has been used in 2 selected songs. In order to measure the impact, surveys have been carried out on different people who are within the range from 18 years to 48 years of age. The responses address issues about the importance of the realism of the virtual voice, the audiovisual content (*videoclip*) that goes together with the music and the preference of the same respondents for a song in which the virtual voice has a more artificial sound. This indicates that the definition of what is considered “real” is linked to the musical context in which the parameters of the virtual singers will be used to complement the musical style.

Keywords: composition, music composition, vocaloid, synthesis, vocal synthesis, music, vst

Índice

Resumen.....	ii
Abstract.....	iii
Índice.....	iv
Índice de figuras.....	v
Introducción.....	1
Planteamiento de la investigación.....	6
Estado del arte.....	7
Marco teórico.....	9
Metodología.....	11
Capítulo 1. La síntesis vocal aplicada en la producción musical de canciones.....	12
1.1. Análisis de la canción “A Realistic Logical Ideologist”.....	22
1.2. Análisis de la canción “DOUKEI”.....	33
Capítulo 2. Impacto de vocaloid en 20 jóvenes limeños entre los 18 y 48 años.....	40
2.1. Análisis de las respuestas sobre cada una de las preguntas.....	45
2.2. Análisis de resultados globales.....	51
Conclusiones.....	57
Referencias bibliográficas.....	60

Índice de figuras

Figura 1. Interfaz del VOCALOID 1.....	15
Figura 2. Interfaz del VOCALOID 2.....	16
Figura 3. Interfaz del VOCALOID 3.....	17
Figura 4. Interfaz del VOCALOID 4.....	18
Figura 5. Interfaz del VOCALOID 5.....	19
Figura 6. Interfaz del VOCALOID 6.....	21
Figura 7. Sesión del VOCALOID 6.....	23
Figura 8. Información del “voicebank” en la sesión del VOCALOID 6.....	24
Figura 9. Información de parámetros de voz en la sesión de VOCALOID 6.....	26
Figura 10. Información de efectos vocales del canal 3.....	29
Figura 11. Información de efectos vocales del canal 4.....	31
Figura 12. Información de efectos vocales del canal 6.....	32
Figura 13. Interfaz general del usuario en VOCALOID 6.....	34
Figura 14. Información de parámetros vocales en VOCALOID 6.....	36
Figura 15. Información de efectos vocales en el canal 3 de VOCALOID 6.....	38
Figura 16. Configuración del efecto de delay de los tracks 4 y 5.....	39
Figura 17. Gráfico circular que presenta a los músicos y no-músicos.....	41
Figura 18. Gráfico circular sobre la familiaridad de entrevistados con cantantes virtuales.....	42
Figura 19. Gráfico de barras sobre la variedad de motores de síntesis vocal que hay.....	44
Figura 20. Gráfico circular sobre la relevancia del realismo vocal.....	45
Figura 21. Gráfico circular sobre la preferencia hacia la canción #2.....	46
Figura 22. Gráfico circular sobre la relevancia de los videoclips.....	47
Figura 23. Gráfico circular sobre el realismo hacia la canción #1.....	48
Figura 24. Gráfico circular sobre la preferencia del estilo musical vocal.....	49
Figura 25. Gráfico circular sobre el reemplazo de cantantes virtuales sobre los reales.....	50

Figura 26. Gráfico circular sobre la relevancia en el campo del *songwriting*.....51

Figura 27. Gráfico circular sobre el rango de edades de los encuestados.....53



Introducción

Justificación

Esta tesis busca la comprensión y apreciación de la música generada por *software* de síntesis vocal identificando sus elementos técnicos como la naturalidad de una pieza con relación a la voz virtual y el factor humano-realista presente en la sección vocal. A pesar de las innovaciones e investigaciones tecnológicas dentro del campo musical en el uso de instrumentos virtuales, el uso de las voces virtuales sintetizadas es un elemento que no se ha investigado desde una perspectiva inmersiva, sino que, más bien, se mantuvo descriptiva sobre lo que se ha ido conociendo y la forma en la que se ha estado utilizando este recurso musical como un instrumento más. También se busca mejorar la creación y producción de música con *software* de síntesis vocal dentro del campo del *songwriting*, es decir, profundizar la aplicación de estos elementos tecnológicos en la composición de canciones. Esto conlleva a la experimentación, tanto sonora como también de estilos, aportando nuevas posibilidades de registro, por ejemplo, y timbre vocal como medio de expresión dentro de una composición original en el marco de una estética perteneciente a una cultura digital.

Motivación

Mi motivación hacia este tema surge a partir de un videojuego de ritmo en el cual las canciones que se escuchan están acompañadas de un “avatar” que baila una coreografía. Este personaje “virtual” ha causado un impacto en los video jugadores, pero también sucede con la audiencia que escucha o que también observa el video del modelo 3D del cantante protagonizando un videoclip en redes sociales como YouTube. Lo que busco es comprender la existencia de una identidad entre la audiencia y un ser virtual. Buscando información sobre el uso de las voces virtuales, al igual que las DAW, existe VOCALOID: una de las tantas “DAWs”, pero usada exclusivamente para tratar voces virtuales. Este campo no se ha explorado mucho desde una perspectiva composicional, por lo que considero que este tema contribuye en el uso de innovaciones tecnológicas dentro de la industria musical. Como

songwriter, el uso que se le atribuya a la voz virtual también es una excelente herramienta que se puede aplicar en futuras composiciones tratándose como un instrumento más.

Relevancia

Actualmente, la tecnología tiene gran importancia en las actividades diarias del ser humano. Dentro del ámbito musical, se han podido apreciar los avances tecnológicos que ha beneficiado a la industria de la música llevando consigo a la creación de nuevos géneros y estilos musicales. La síntesis vocal en la música hace referencia al uso de los *softwares* que permiten “emular” el canto humano y que circuló en el mercado desde el año 2008, cuando Yamaha Corporation anuncia la salida al mercado de VOCALOID, un *software* que permite utilizar “cantantes virtuales” a través de un ordenador vía MIDI. VOCALOID se establece como una fuente de tipo sonora, ya que su función es producir el sonido por medio de un ordenador. Sin embargo, también es posible catalogarla como una fuente de tipo audiovisual, ya que existen varios videoclips que engloban a este mismo término, pues también ha habido una transición entre lo que *inicialmente* es hasta lo que *actualmente* es dentro de un campo sociocultural sobre cómo el uso de cantantes virtuales ofrece nuevas posibilidades para establecer formas de expresión entre el artista y el oyente bajo el análisis estético de los recursos expresivos de la música manifestado en la composición y producción de canciones.

La importancia de profundizar en el estudio de la síntesis vocal a través de los años consiste en revisar detalladamente el uso que los artistas les otorgaban a las voces virtuales como una herramienta orientada a la actividad del canto, centrándose en la conexión entre el compositor y la audiencia. La música hasta el día de hoy es una de las tantas formas en las que el ser humano consigue expresar numerosas emociones por medio de los sonidos.

Progresivamente, la música y la tecnología han permitido novedosas formas de creación de contenido como la utilización de instrumentos virtuales que promueven un sinnúmero de posibilidades en el proceso creativo que han ido evolucionando a lo largo del tiempo (Wilkinson, 2003, p. 1). Una de estas innovaciones es la síntesis vocal por medio de un

software de computadora siendo un recurso que causó cierta polémica en su tiempo por lo contraproducente que podría llegar a ser dentro del campo musical. Esto se debe a que se temía que, en algún momento, sea un reemplazo frente a los cantantes reales si el usuario dispone del conocimiento requerido. Sin embargo, las consecuencias fueron diferentes en aspectos sociales, seguido de diversas opiniones con respecto a todo lo que engloba. De los varios sintetizadores de voz que existen en el mercado, voy a hacer referencia al *software* producido por Yamaha Corporation llamado VOCALOID, que permite emular el estilo de canto de una persona a través de fonemas que se encuentran almacenados en bancos de voz (*voicebanks*) interpretados por una base de datos del mismo *software*. Esto trajo consigo al inicio de una posible “era virtual” en el área musical. Numerosas personas cuestionaban las posibles consecuencias que podría causar, generando un grupo que estaba a favor y en contra. Por ejemplo, se menciona que existe el temor de adquirir el desarrollo de tendencias antisociales a causa de una dependencia de los recursos tecnológicos con las relaciones sociales de los músicos (Collins, 2011, p. 1). Otros mencionan que es una nueva forma de entretenimiento que impulsa a la globalización y una nueva forma de producción en la música por un intercambio cultural (Yin, 2018, p. 52).

Cuando salió VOCALOID al mercado, los primeros bancos de voz que salieron a la venta carecían de una imagen propia que los caracterizara. En el artículo “The dB in the .db: Vocaloid Software as Posthuman Instrument” de la revista *Popular Music and Society*, Bell afirma lo siguiente: “the first VOCALOID products released were intended to replace live studio back-up singers but were not widely adopted for that purpose” (2016, p. 6). Como lo señala la autora, en sus inicios este *software* no estaba pensado inicialmente para el público en general: la intención con este programa era emplearlos como un reemplazo o apoyo para cantantes que tenían presentaciones en vivo. Por eso, los primeros bancos de voz se vendían sin una imagen que sea propia de ello. Sin embargo, no se generaban buenas ventas debido a

la ineficacia de los mismos bancos de voz en ese momento, que eran generados virtualmente desde cero. Así como los artistas tienen una imagen personal que los hace diferentes al resto de personas, se optó por crear una imagen propia a los bancos de voz para generar mayor popularidad en su próxima entrega. De esta manera, en el año 2007, la empresa Crypton Future Media que colaboraba con Yamaha en el proyecto, lanzó una nueva versión del *software* denominado VOCALOID 2, en la cual hicieron mejoras en la calidad vocal, la dicción y ornamentos en la voz. Asimismo, se lanzó un nuevo banco de voz en el año 2007 que causó una impresión en los usuarios de dicho programa: Hatsune Miku, que significa “el sonido del futuro”.

Cuando Yamaha Corporation comenzó a vender su *software* al público, utilizó una etiqueta o eslogan en la que sugiere una idea al consumidor: “Human beings are still singing? We are in the digital age! ... It’s time for you to join us!”¹(2016). Con este enunciado, se plantea el inicio de una “era virtual” al usar cantantes virtuales planteando dos posturas: la conservadora y la progresista. Algunos usuarios plantean que “existe una posibilidad de adquirir el desarrollo de tendencias antisociales a causa de una dependencia de estos recursos tecnológicos” (Collins, 2011, p. 1). Otros solo mencionan que es una nueva forma de entretenimiento que impulsa a la globalización y una nueva forma de producción en la música por un intercambio cultural (Yin, 2018, p. 52). El mundo en el que se vive depende de la tecnología que ha estado en constante desarrollo y lo mismo va aplicado a la música: antiguamente se escribía a mano en un pentagrama durante la época clásica y con el tiempo surgen *softwares* de notación musical hasta que, con la llegada de las Digital Audio Workstation es más práctico simular hasta una orquesta con unos cuantos instrumentos virtuales en el ordenador que terminan de alguna forma omitiendo el proceso de grabación de instrumentos reales (Klein, 2016, p. 23). Ciertamente el uso que se le ha dado a los

¹ La traducción de este eslogan es: “¿Los humanos aún están cantando?;Estamos en la era digital! ... ¡Es hora de que te unas!

instrumentos virtuales está más visto para crear un demo de alguna pieza musical y eso es debido a que tiene un objetivo dentro de esta área.

Sin embargo, esto lleva a tomar lo que mencionó Yin previamente y consiste en la aceptación de la llegada y creación de canciones con cantantes virtuales como un recurso más. Los primeros productos relacionados a VOCALOID como actualizaciones del *software* en sí o la venta de nuevos bancos de voz tenían la intención de reemplazar a los coristas que suelen hacer el *background voice* o también llamado “segunda voz” pero no fueron ampliamente adoptados para ese propósito (Bell, 2016, p. 6). Describirlo como un “reemplazo” puede sonar chocante para cierto grupo de personas hasta el punto en llegar a ser considerado antiético porque es quitarle protagonismo: quitar la importancia de tener un coro vocal “real” y optar por usar una máquina. Kenmochi menciona en un artículo que, al adjuntar las identidades de los cantantes reales al producto y al construir personajes para bibliotecas de muestras para VOCALOID, marcan un espacio entre “lo real” y lo “imaginario” (2012, p. 5387). Es decir, el objetivo de las voces virtuales es simular ese carácter “humano” en lugar de imitarlo, ya que usar un ordenador que simule la habilidad para cantar al punto de parecerse mucho a la voz humana, por más limpia que pueda ser la entonación, no va a pasar más allá de eso: una imitación.

Planteamiento de la investigación

Pregunta de investigación

¿De qué manera el uso del sintetizador VOCALOID se consolida como una herramienta para la composición de canciones y cuál es su impacto en el público de la ciudad de Lima entre los 18 a 48 años de edad?

Preguntas secundarias

- ¿De qué manera el uso del sintetizador VOCALOID aporta en la construcción de un estilo musical propio?
- ¿Cómo la experimentación de estilos musicales con voces sintetizadas ha tenido influencia en la población seleccionada?

Objetivo principal

Comprender los elementos que, por medio de la síntesis vocal, se consolidan como una herramienta para la composición de canciones como medio de expresión musical

Objetivos secundarios

- Evidenciar la importancia del uso del sintetizador VOCALOID en la construcción de un estilo musical propio
- Analizar cómo la experimentación de la síntesis vocal impactó en la población seleccionada

Estado del arte

En el artículo titulado “Feigning Humanity: Virtual Instruments, Simulation and Performativity”, publicado en la revista *Journal of the International Association for the Study of Popular Music*, Eve Klein menciona las formas en que los *softwares* de instrumentos virtuales (VI) simulan la acústica de la interpretación humana. Se habla de dos estudios de caso: los instrumentos de orquesta virtual (VOI) e instrumentos de canto virtual (VSI) considerando en cómo su diseño e implementación busca expresar la interpretación musical humana mediante la adopción de naturalidad, tono, dinámica, articulación, ambiente y otras limitaciones impuestas por la relación física entre el ejecutante y el instrumento (2016). Un detalle que resalta el autor es que cataloga las voces virtuales como un instrumento adicional, además que sirve como un medio por el cual se transmiten los gestos de la expresividad musical. Este enunciado entra en cuestión ya que menciona que algunos oyentes cuestionan la interpretación humana por revelar la imitación. Sin embargo, esta “imitación” hace referencia a un “legato imperfecto” o incluso a un *glitch*² en el texto en el momento de renderizar los fonemas. Tratar este tipo de detalles destaca el proceso de producción y, consecuentemente, las nociones del “factor humano” y la expresividad en la interpretación musical. Estos conceptos de “humanización” y caracteres humanos presentes en las voces virtuales son parámetros que también menciona Cuevas en su tesis *Reportaje audiovisual VOCALOID: los cantantes del futuro*. Estas características son relevantes en la simulación de la interpretación humana mediante el uso de instrumentos virtuales.

Dentro del campo de investigación que estudia el conjunto de elementos que abarcan las “bandas virtuales”, se señala que la era virtual significaría una nueva frontera en el área de los estudios musicales en la cual se investigan sobre los elementos que conforman las “bandas” o “idols virtuales”. El artículo “Living in the Inmaterial World: Holograms and

² Un pequeño problema o falla que impide que algo tenga éxito o funcione tan bien como debería ser.

Spirituality in Recent Popular Music” afirma que la cultura de lo digital ha sido testigo del cambio que hay desde el paradigma de la “realidad virtual” y el paradigma de “lo real” (McLeod, 2016). Además, analiza las actuaciones de los artistas que engloban el mundo virtual en el rubro de la música popular. El autor también afirma que las actuaciones holográficas son producto del rasgo “humano” en el que hay una interacción entre el holograma y los músicos. Esto crea progresivamente la “conciencia humana colectiva” que es el nexo entre la figura del intérprete con el oyente. Este concepto también es mencionado por Kenmochi en un artículo, en el cual afirma que, al adjuntar las identidades de los cantantes reales al producto y al construir personajes para bibliotecas de muestras para Vocaloid marcan un espacio entre “lo real” y lo “imaginario” (2012, p. 5387). Esto se refleja en el eslogan del *software* de Yamaha Corporation: “Humans are still singing? We’re in the digital age! It’s time for you to join us!” (2016).

Finalmente, el artículo “The Vocaloid Phenomenon: A Glimpse into the Future of Songwriting, Community-created Content, Art, and Humanity” trata el tema de cómo los medios de producción musical han ido evolucionando y la influencia que tiene específicamente en el campo de lo que constituye al songwriting. El autor explica cómo la intención del *software* “Vocaloid” estaba orientado para los *songwriters* que no pueden cantar sus propias canciones (Roseboro, 2019). Sin embargo, también detalla que la adquisición de los *voicebanks*, por el cual el sintetizador toma las muestras de voz en una base de datos, eran vendidos con una ilustración de la artista virtual que lo representaba. Las personas, al crear sus propias canciones por medio del *software*, han provocado que se popularice en las redes sus creaciones junto a diversas animaciones sobre el/la “cantante virtual”. Debido al estilo único que posee, la comunidad que engloba a la música creada por la síntesis vocal lo cataloga como un “nuevo estilo” musical, donde se manifiesta una libertad expresiva durante el proceso de creación en el que el cantante virtual expresa “lo humano” del autor.

Marco teórico

Esta investigación se enfoca en cómo el *software* de síntesis vocal ‘VOCALOID’ se considera como una herramienta tecnológica musical en la cual es tomado como un “instrumento” más para la composición de canciones.

En el capítulo 22, “New Means of Creation”, del libro *Basic Music Technology: An Introduction* escrito por Guerino Mazzola y otros, se señala que las formas de composición musical de la actualidad se han vuelto más asequibles para el común de las personas, las cuales no posean los mismos conocimientos de teoría musical que un músico formado, y que esto convierte a la creación musical en un medio de fácil acceso para la expresión humana. Por lo tanto, el autor señala la importancia de comprender los mecanismos por los que se produce la música a través de los medios tecnológicos modernos (2018, p. 177). Comprender estos parámetros que son parte del proceso de creación musical permite que el compositor o arreglista potencie sus conocimientos sobre los recursos que dispone para llevar la producción en una dirección específica a nivel de producto. De igual manera, en la tesis *The Vocaloid phenomenon: A glimpse into the future of songwriting, community-created content, art, and humanity*, escrita por Bronson Roseboro, dicta que los elementos que abarcan la producción musical implican el tener que centrarse en los arreglos, la orquestación, el proceso de mezcla y, finalmente, la masterización (2019, p. 11). Todos estos parámetros que conforman el proceso creativo son elementos parte de la composición musical.

En el artículo “Singer-songwriter meets music production and studio technology”, redactado y publicado por Richard James Burgess y Rob Toulson en la revista *The Singer Songwriter Handbook*, se señala que los cantautores suelen ser los que maquetan su propia propuesta de arreglo musical añadiendo secciones de instrumentos que ellos mismos no podrían tocar con soltura (2017, p. 2). Es decir, adoptar la tecnología como un recurso adicional es esencial para hacer realidad su visión creativa personal. Asimismo, el mismo

autor señala que el estudio y material que permita aportar en todo el proceso de creación y producción de la obra, es considerado una herramienta que aporta en el proceso, convirtiendo el 'estudio' a tratarse como un 'instrumento musical' (p. 6). Cabe mencionar que el estudio como "instrumento musical" generalmente se refiere al acto de ingresar al estudio o comenzar a grabar antes de que se hayan realizado composiciones o arreglos musicales detallados, lo que significa que el proceso creativo está fuertemente influenciado por la tecnología de la grabación dentro del espacio del estudio.

En el artículo "La música y su interpretación como vehículo de expresión y comunicación" realizado por la violinista Danuta Glowacka en la revista *Comunicar*, se destaca la intención comunicativa de la música, en la cual resaltan los valores emocionales y expresivos que los músicos sienten y perciben a través de la música que interpretan (2004, p. 1). Esto afirma que la función comunicadora siempre ha estado presente a lo largo de la historia de la música, desempeñando su accionar como intérpretes-comunicadores y como creadores-expresivos. La música siempre ha sido un medio de expresión y su relevancia permite crear un vínculo entre el 'emisor' (cantante o intérprete) y el 'receptor' (audiencia), haciendo énfasis en la intencionalidad del proceso creativo.

Metodología

El tipo de investigación es de enfoque cualitativo. Se analizaron dos canciones y se observó cómo ha sido empleado el sintetizador VOCALOID en la composición y producción musical de ambas canciones replicando en el mismo *software* el uso dado a las voces virtuales. Asimismo, se recolectaron datos mediante unas encuestas estructuradas, preguntando a las personas sobre la percepción que tienen de las dos canciones en las que se ha empleado el *software* VOCALOID junto a este tipo de música creada con ‘cantantes virtuales’ para luego contrastar la información de la encuesta con el análisis musical realizado, midiendo el impacto que causa este tipo de música sobre los oyentes.



Capítulo 1. La síntesis vocal aplicada en la producción musical de canciones

A lo largo del tiempo, se ha podido apreciar diversos avances tecnológicos que en diversas industrias, dentro del campo de la industria musical, la tecnología ha llevado a la creación de nuevos géneros y estilos musicales a través de la experimentación sonora y el uso de nuevos recursos. Desde los 80's, la aparición de la tecnología MIDI (Musical Instrument Digital Interface) se transformó en una herramienta importante, marcando el inicio de una nueva era en la creación musical, siendo esta la era de la música digital o computarizada. En un artículo de BBC News, se señalaba precisamente que ninguna persona imaginaba que implementar una interfaz digital con instrumentos musicales era el primer paso para que los procesos de creación y experimentación sonora se encuentren más al alcance de las personas, lo que supuso un cambio enorme (2012). Una de las ramas de la experimentación sonora consiste en la síntesis vocal, que se define como la imitación del habla humano a través de la creación de una voz artificial, llamándolo “banco de voz” o *voicebank*.

La síntesis vocal ha sido considerada como un instrumento desde sus orígenes debido a que sólo tenía la función de emular el habla humana. Cuando surge VOCALOID como un *software*, permite usar la síntesis vocal con la tecnología MIDI para simular el canto humano a través de un proceso de síntesis concatenativa, que consiste en la unión de segmentos de voz grabadas o emuladas. De esta manera, la síntesis vocal comenzó a ser una herramienta útil para la composición de canciones, aunque en sus inicios no haya sido pensado para el público en general, pues la intención con este programa era emplearse como un reemplazo o apoyo para los cantantes que tenían presentaciones en vivo. Según Bell, los primeros productos relacionados al *software* VOCALOID se vendieron con la intención de servir como un reemplazo a los cantantes “reales” que realizaban los coros de apoyo pero que no fue utilizado con ese propósito (2016, p. 6). En contraste, Isernte menciona que en el canto coral no existen muchas herramientas de apoyo para el cantante sin la posibilidad de tener

intérpretes “reales” llevando a cabo la elaboración de partituras como parte del repertorio musical para mezclar las voces armando el ensamble en la que es muy importante tener un control de la expresividad simulada en los parámetros del *software* (2014, p. 4), optimizándose desde VOCALOID 2 hasta las versiones más recientes.

El primer motor de síntesis vocal para el canto fue VOCALOID 1, desarrollado por Yamaha Corporation junto a Zero-G Limited en el desarrollo de los primeros *voicebanks* de *soul*, LEÓN y LOLA, cuyo lanzamiento conjunto se dió en el año 2004. Los parámetros que se manejan en esta versión del *software* son más primitivos y sin la misma versatilidad y precisión que en sus versiones más recientes en los efectos que se aplican a algunas voces. Hay elementos visuales en la interfaz que dificultan la edición, como la paleta de dinámicas de expresión que representan valores fijos y mecánicos, la ausencia de poder visualizar los canales de audio en una línea de tiempo como en otras DAW, la falta de poder importar un audio en formato WAV para trabajar las voces con la pista musical de fondo y los archivos de guardado como “VOCALOID MIDI File”. Cuando se usa como un VSTi accesible desde una DAW, el *plug-in* VST omite el editor y envía directamente esos mensajes al motor de síntesis, encargándose de interpretar los datos para reproducirlos. En el caso de los *voicebanks* que acompañaron al *software* en su lanzamiento, también presentaban algunos problemas de pronunciación al procesar ciertos sonidos. Es por ello que puede afirmarse que esta versión del *software* era vista como un prototipo de lo que posteriormente pasará a consolidarse en el futuro.

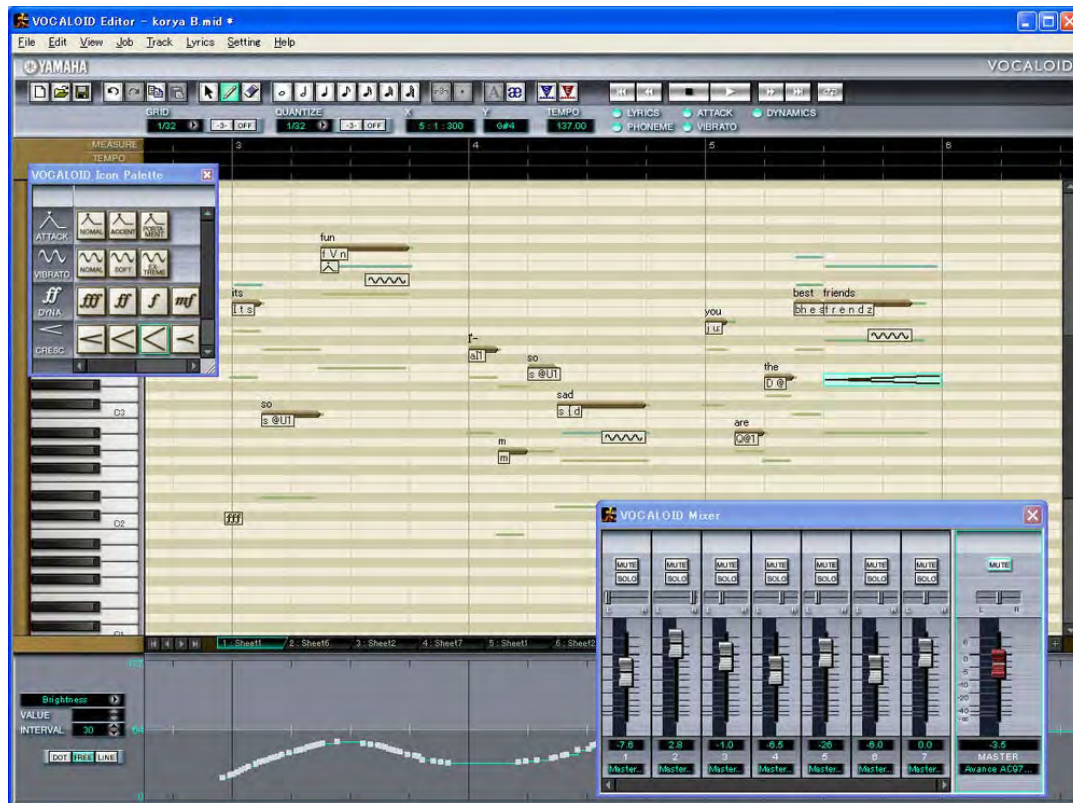
Existen parámetros que permiten editar muchas propiedades de las voces virtuales a usarse. El parámetro *Brightness* (Brillo), permite alterar la voz virtual enfatizando las frecuencias más altas, es decir, aumentar este parámetro ocasionaría una voz más enérgica y disminuirlo, una voz más dulce. *Clearness* (limpieza) determina la nitidez de la voz. Aumentar este parámetro genera una voz más “aguda”, mientras que disminuirlo causa una

voz más “apagada”. El parámetro *Gender Factor* (factor de género) va a determinar el tono que adquiere la voz. Aumentar este parámetro genera una voz llevándola a que suene más “femenina” y disminuirlo genera una voz más “masculina” y gruesa. *Harmonics* (armónicos) consiste en la cantidad de armónicos que se escucharán en la voz. Este parámetro no estará presente desde la tercera entrega del *software*. El parámetro *Noise* (ruido) está anexado al parámetro previamente mencionado. Éste se encarga de controlar el nivel de la respiración simulada por el *software* al igual que la resonancia generada por la misma voz como consecuencia de los armónicos (no es un REVERB). *Note Velocity* controla el volumen y acento individual de cada nota. Este parámetro es el equivalente en lo que posteriormente se llamará VELOCITY y será de gran uso al importar archivos MIDI en VOCALOID.

Resonances (resonancia) consistía en un conjunto de opciones que permitían editar los fonemas específicamente asignados a una nota. Editar el fonema afecta a la pronunciación de las vocales y, por consiguiente, al timbre vocal. Este es un parámetro que posteriormente no va a estar presente en el *software*. *Singer* (cantante) es una opción que permite cambiar el cantante que se utiliza para reproducir las notas. El *Pitch Bend* edita la “curvatura” del tono entre una nota y otra. *Pitch Bend Sensitivity* está anexado al parámetro anterior y permite ajustar el valor del rango del *Pitch Bend*. Finalmente, el parámetro *Portmento Timing* se utiliza para cambiar el “tempo” de las notas. Este parámetro está ligado al estilo de canto que adopta la voz virtual.

Figura 1

Interfaz del VOCALOID 1



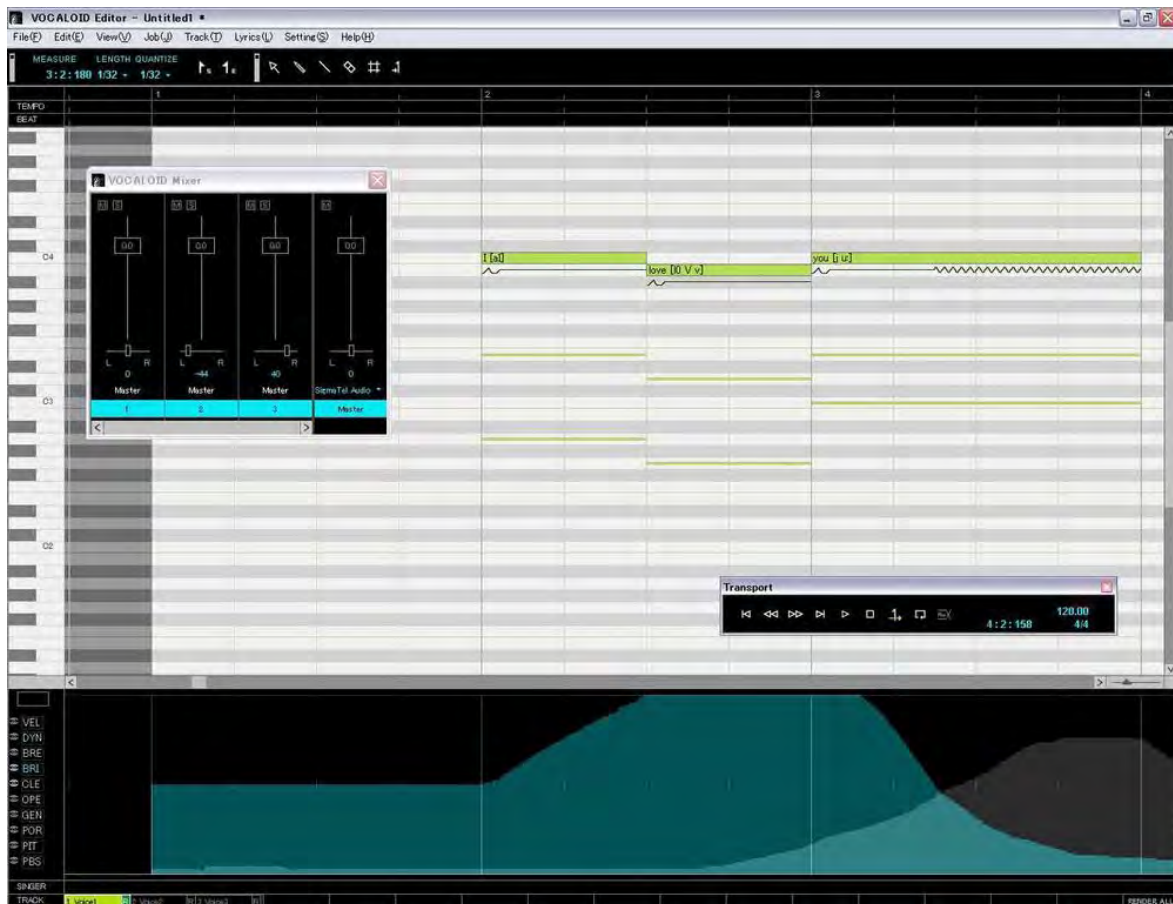
Nota. Se muestran las dinámicas de expresión, notas con fonemas importados de un archivo MIDI y la ventana del mixer.

En el año 2007, se comercializa el motor de VOCALOID 2 presentando múltiples mejoras en la calidad vocal, la dicción y ornamentos dentro del editor. Esta versión no posee retro compatibilidad, por lo que los archivos de guardado de VOCALOID 1 no son compatibles en VOCALOID 2. Los archivos de guardado de esta versión son en formato “VSQ” En cuanto a la interfaz de usuario, presenta una mayor variedad de ajustes que pueden agregarse para el manejo de las voces virtuales como la mejora en la pronunciación de ciertos fonemas. Asimismo, se lanzó un nuevo banco de voz en ese mismo año que causó una impresión en los usuarios de dicho programa: Hatsune Miku, que significa “el sonido del futuro”. Una característica de este *voicebank* consiste en que posee un amplio rango vocal sin evidenciar su “artificialidad”. Las autoras Jorgensen, Vitting-Seerup y Wallevik, comentan que los *fans* pueden comprar y componer canciones con el *voicebank* de Hatsune Miku de la

misma forma como si escribiesen para un instrumento virtual de un *software* de producción musical (2017, p. 2).

Figura 2

Interfaz del VOCALOID 2



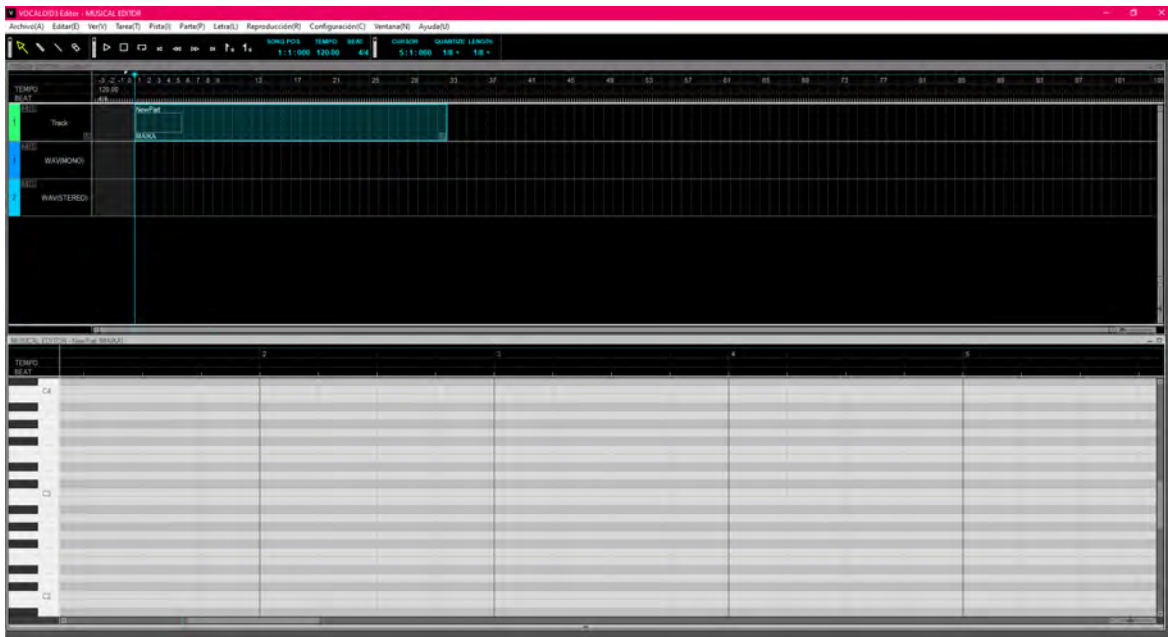
Nota. Se muestran los parámetros disponibles, notas insertadas con los fonemas y la ventana del mixer.

La tercera entrega del *software* de síntesis vocal, VOCALOID 3, presenta consigo los mismos parámetros de su predecesor con la inclusión de nuevos *voicebanks* y mejoras del motor de síntesis. El renderizado de las voces ha recibido una mejora ya que dispone de un botón que “unifica” los fonemas para que reproduzca un sonido más orgánico. Además, a partir de esta entrega el *plug-in* VST se adquiere por separado y se trabaja por defecto con la versión *stand alone*. A diferencia de sus predecesores, presenta una mejor interfaz de usuario permitiendo importar un archivo WAV para trabajar con la pista de fondo al igual de poseer retrocompatibilidad. También dispone de mayor cantidad de *job plugins*, que corresponden a

complementos para trabajar con el *software* y ofrecer mayores recursos en el uso del motor de síntesis. Por ejemplo, se agrega el complemento de “VOCALOID 3 API”, en el cual se permite grabar una muestra de voz en tiempo real para que el usuario pueda manipular su propia voz para “crear” primitivamente un *voicebank* de apoyo.

Figura 3

Interfaz del VOCALOID 3

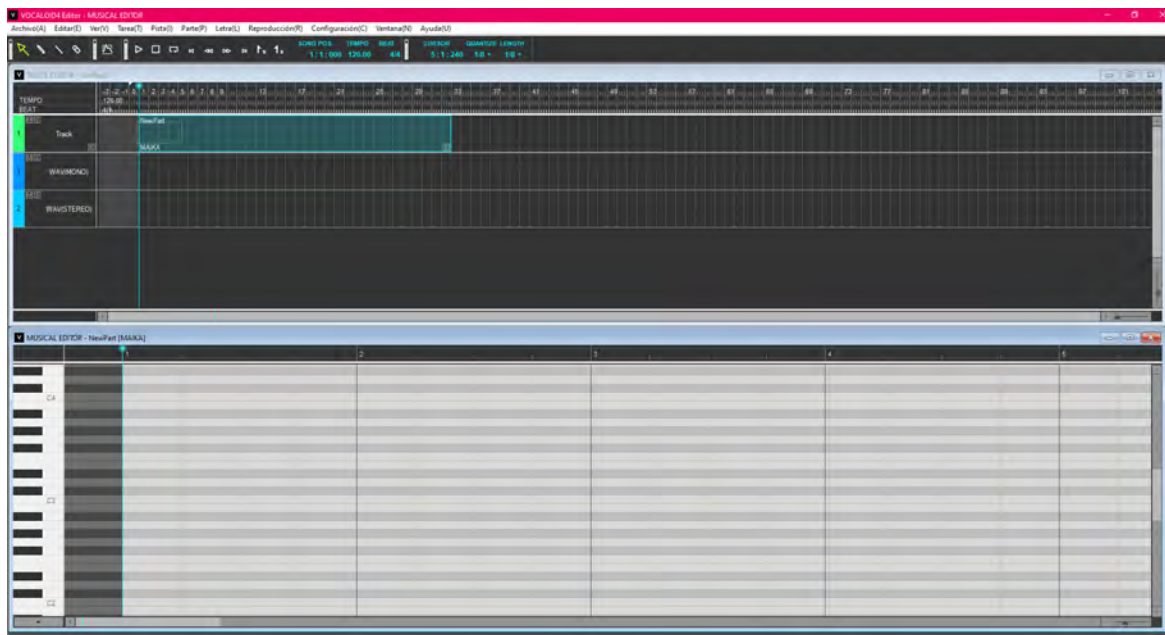


Nota. Interfaz del usuario al abrir una nueva sesión.

La cuarta entrega, VOCALOID 4, ofrece una interfaz similar si se compara con su versión previa al igual que las funciones y *job plugins* que el programa ofrece. Una de las funciones nuevas es el parámetro de *growl*, el *Pitch Rendering*, *Pitch Snap* y el *cross-synthesis*. El nuevo parámetro *growl* permite que la voz virtual emule la técnica vocal asociada a dicha sonoridad de “gruñido” en las notas. La opción de *Pitch Rendering* permite visualizar graficamente el estado actual del tono aplicado a la secuencia de notas insertadas en el editor. La opción *Pitch Snap* ofrece el uso de voces con sonoridad robótica ya que desactiva la curvatura del tono de las notas. Finalmente, la opción de *cross-synthesis* permite cambiar entre dos *voicebanks* dentro de una misma canción. Esto solo es posible si los *voicebanks* seleccionados son compatibles con esta nueva función.

Figura 4

Interfaz del VOCALOID 4



Nota. Interfaz del usuario al abrir una nueva sesión.

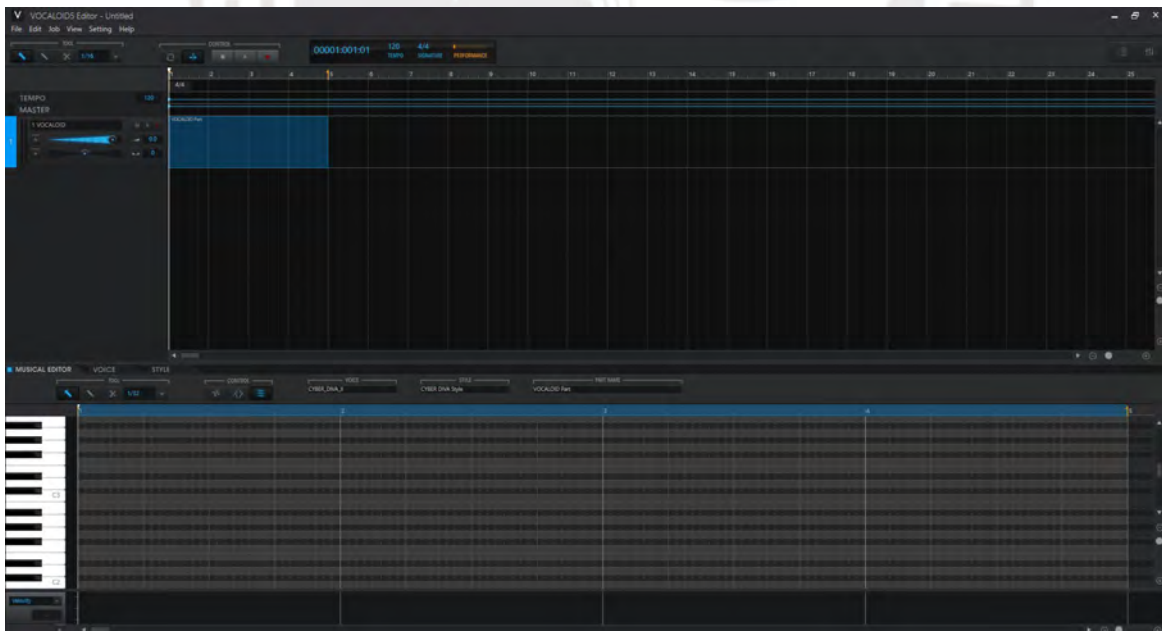
El lanzamiento al mercado del motor de VOCALOID 5 supuso un cambio total en la interfaz de usuario. Ahora el proceso de renderización es siempre de forma automática al momento de realizar cambios en la melodía. A diferencia de los *Job Plugins* que se muestran desde VOCALOID 1, algunos de los complementos se incluyen dentro del mismo *software*. Además, incluye varias configuraciones preestablecidas para distintos tipos de canto y frases. Soporta hasta 32 tracks, siendo el doble de canales disponibles a usar en comparación con las entregas anteriores, al igual que la lectura de audios en formato WAV. Se vuelve a implementar una versión del *software* para usarse como un VST en otras DAW, pero en su versión beta, esto porque el motor de síntesis tiene un apartado de funciones adicionales en contraste con sus predecesores. Esta versión del *software* guarda los archivos con la extensión VPR³, que compone a un conjunto de elementos como audio, ecualización y demás elementos de edición compresos en un solo archivo. Un dato que lo diferencia de las entregas

³ Formato de archivo de guardado para VOCALOID 5 y VOCALOID 6. Contiene un registro de las notas musicales, letras y voces sintetizadas que contiene el proyecto.

anteriores es que al usarse como VST, se puede guardar la secuencia no solo como secuencia MIDI, sino también como un archivo en formato VPR para poder trabajarse más detalladamente en su versión *stand-alone*, que también permite abrir archivos en formatos VSQ y VSQX. Entre los puntos negativos, parámetros de modulación de las voces virtuales como el *cross-synthesis* ya no estará disponible, al igual que los *voicebanks* que se lanzaron para VOCALOID 2 que no son compatibles con la entrega de VOCALOID 5 en adelante. Esta entrega fue la que ha tenido cambios mayores a nivel general y no ha sido bien optimizada pese a recibir actualizaciones. Una evidencia de este desperfecto se presenta en la parte superior que dice “PERFORMANCE”. Este apartado funciona como un indicador sobre los recursos que consume el *software* en el sistema operativo. Cuanto mayores recursos del sistema requiera el uso del *software*, la barra irá en aumento.

Figura 5

Interfaz del VOCALOID 5



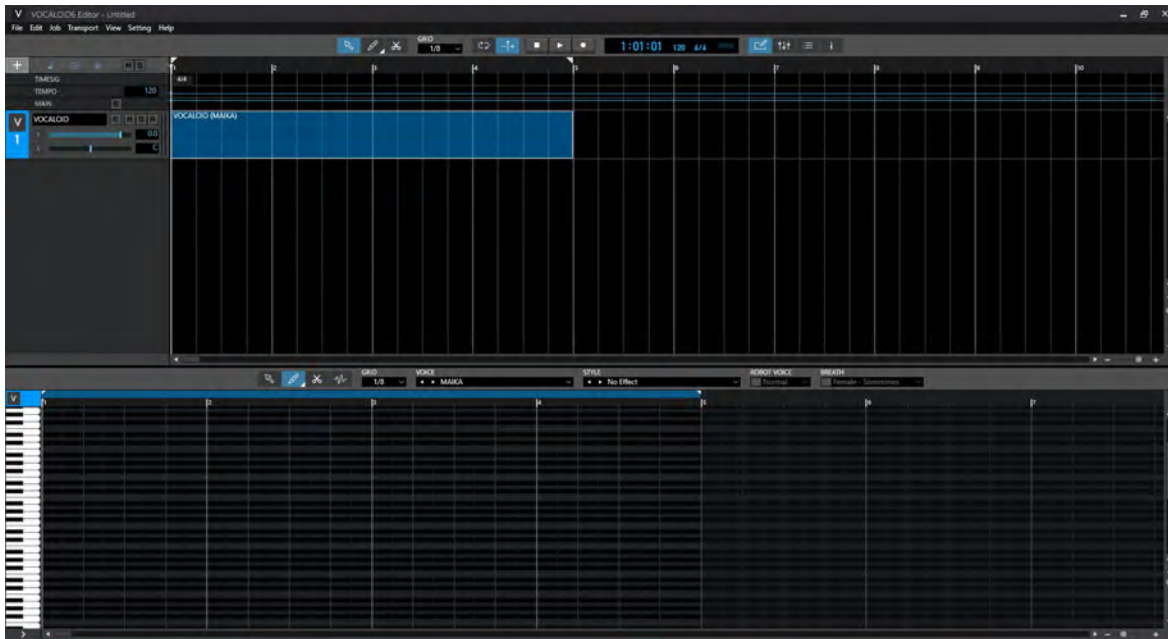
Nota. Interfaz del usuario al abrir una nueva sesión.

Con el pasar de los años, el *software* en cuestión ha pasado por varias revisiones y actualizaciones con la finalidad de mejorar su interfaz de usuario, rendimiento y la

compatibilidad con otras DAW. VOCALOID 6 es el motor de síntesis más reciente de Yamaha y presenta mejoras considerables con respecto a su entrega anterior. Al igual que sus predecesores, está al alcance de cualquier usuario que disponga de una computadora. Los parámetros que se muestran en pantalla tienen un parecido con los de alguna otra DAW, como Logic Pro X, Garage Band, Ableton, Reaper y similares. Además de una interfaz similar en el aspecto visual, se ha mejorado la facilidad del uso con otras DAW como *plugin* VST en las versiones de Windows y AU en las versiones para Mac. Este *plugin* ahora trabaja con una extensión llamada “Audio Random Access” o abreviado como ARA. Dentro de VOCALOID 6, este *plugin* recibe el nombre de “VOCALOID Bridge” y permite actuar como un nexo e intercambiar gran cantidad de información entre el motor de síntesis VOCALOID 6 y la DAW. De esta forma, se podrá trabajar con VOCALOID 6 en un canal de instrumento dentro de la DAW. Otro nuevo elemento que se añade a esta versión es el uso de la inteligencia artificial dentro del proceso de síntesis llamado VOCALOID:AI. Este recurso permite a los usuarios ingresar notas y letras y el *software* no solo rastreará los datos, sino que también determinará los matices que se realizarán en las notas, por ejemplo, seleccionando timbres, conectando notas, aplicando vibratos, etc. Es decir, VOCALOID:AI es diferente de sus predecesores porque es el propio sintetizador el que puede hacer su propia expresión de maneras en la que el creador nunca hubiera pensado. Conociendo esta función que ofrece esta entrega del *software*, también deriva una función con el nombre de “VOCALO CHANGER”, en la cual el usuario puede importar en WAV una muestra de voz grabada en tiempo real para que el motor de síntesis lo replique como un *voicebank* propio. Este elemento es similar al VOCALOID 3 API que se ha desarrollado mejor para esta entrega.

Figura 6

Interfaz del VOCALOID 6



Nota. Interfaz del usuario al abrir una nueva sesión.

En todas las entregas de VOCALOID desde su primera versión hasta la última en mención, ha estado sirviendo como un complemento para la creación de líneas melódicas en cuanto al apartado vocal. Tanto la versión *standalone* como la versión VST/AU, han estado ofreciendo facilidades para servir como una herramienta musical para el compositor. La versión VST permite usar los elementos básicos de VOCALOID dentro de la DAW mientras que la versión *standalone* permite importar y exportar los archivos de audio que se usen, al igual que los *voicebanks* que se emplean para reproducir el canto, exportarse en formato WAV para su uso posterior en una DAW. El *software* también permite importar un archivo MIDI de cualquier programa de notación musical con información de texto que haya sido agregado a una sección vocal. Por ejemplo en el caso del *software* Sibelius, uno puede agregar una sección de instrumento. Si a esa sección se le agrega letra con la opción “lyric input”, esa letra que uno agrega se guarda en el archivo MIDI que se exporta, transfiriéndose en VOCALOID con las sílabas correctamente asignadas. Se puede ver claramente que los

productores enfocados en el uso del *software* han experimentado los distintos usos que se le pueden dar al usar una voz virtual (Bell, 2016, p. 227).

1.1. Análisis de la canción “A Realistic Logical Ideologist”

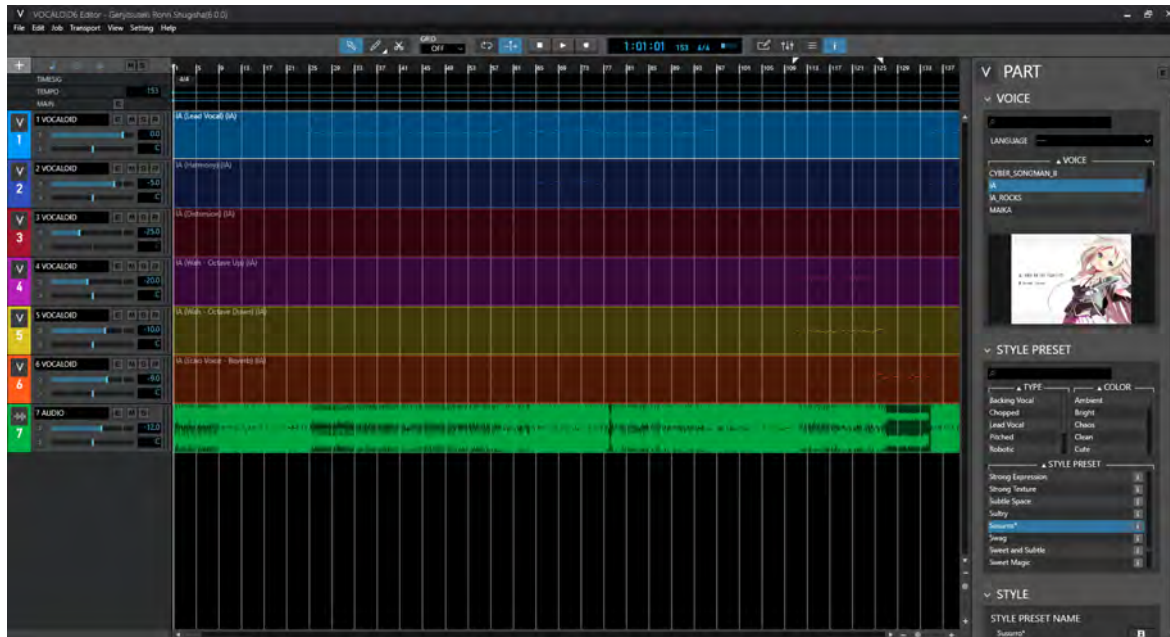
Esta canción fue originalmente compuesta por un productor de VOCALOID que se hace llamar por el pseudónimo de Goboumen. El tema fue producido por el motor de VOCALOID 3 con el *voicebank* de IA y publicado el 8 de marzo del 2012 principalmente en redes como *YouTube* y *Nico Nico Douga* con una duración total de 4 minutos y 29 segundos, siendo parte del álbum “IA/01 –BIRTH-” (2012), “IA/01 –BIRTH- MD” (2012) y finalmente del álbum “IA SUPER BEST –THE CREATORS-” (2022). Esta canción también ha tenido una aparición en el videojuego rítmico “IA/VT –COLORFUL-” (2015) para la consola portátil *Play Station Vita*.

La canción pertenece al género J-rock o también llamado ‘rock japonés’, que es la música rock de Japón influenciada por el rock estadounidense y británico. La música rock japonesa se ha convertido en un culto en todo el mundo, siendo ampliamente conocida en Asia y ha sobrevivido a través de décadas compitiendo con su estilo local derivado contemporáneo J-pop.

Para llevar a cabo el análisis de la canción a detalle, se ha replicado el audio en el *software* VOCALOID 6 en su versión *standalone* para un mejor entendimiento sobre lo que está pasando con el manejo de las voces virtuales, los parámetros que permite el propio *software* para que las voces tengan el tono que se busca y los efectos de sonido que se les puede aplicar para contextualizar mejor el carácter que se le da a la canción. En la Figura 7 se puede observar la información del cantante virtual utilizado, el estilo de canto del *voicebank* y los canales que conforman la sesión siendo 7 canales en total: 6 canales VOCALOID con su respectivo editor de notas y 1 canal de audio.

Figura 7

Sesión del VOCALOID 6



Nota. Menú lateral derecho muestra información de la voz que conforma el track seleccionado.

La canción tiene como único banco de voz a una VOCALOID con el nombre de IA, creada por 1st PLACE Co., Ltd. originalmente para el motor de VOCALOID 3. Este *voicebank* tiene la característica de tener un mayor parecido con la proveedora de voz Lia, vocalista japonesa conocida como “la cantante de la voz de cristal”. A diferencia de otros *voicebanks* que buscan tomar las muestras de voz para crear un timbre propio, la voz virtual de IA busca parecerse a su proveedora de voz. Este detalle es importante ya que lo que se busca es que el producto musical suene natural. Para conseguir esto, no es suficiente con solo insertar las notas en el editor, sino también realizar modificaciones a la línea melódica para que la voz responda poco a poco simulando alguna técnica o acentuación vocal que uno desee. En el texto insertado en el editor, se observa en la Figura 8 que dice “nyuu” en el compás 25 del editor. Lo normal es que esa palabra se mantenga en un Eb5 pero se agregó ese ornamento a la voz partiendo de un Db5. Igualmente, se utilizó el parámetro VELOCITY para articular la acentuación de cada nota insertada en relación a la pronunciación de los

preestablecidos del estilo de canto, el *software* ofrece al usuario configuraciones “por defecto” para que acceda rápidamente a una sonoridad preestablecida. Finalmente, se muestra el editor del estilo de canto, con información más detallada y personalizable para moldear el tono del *voicebank*.

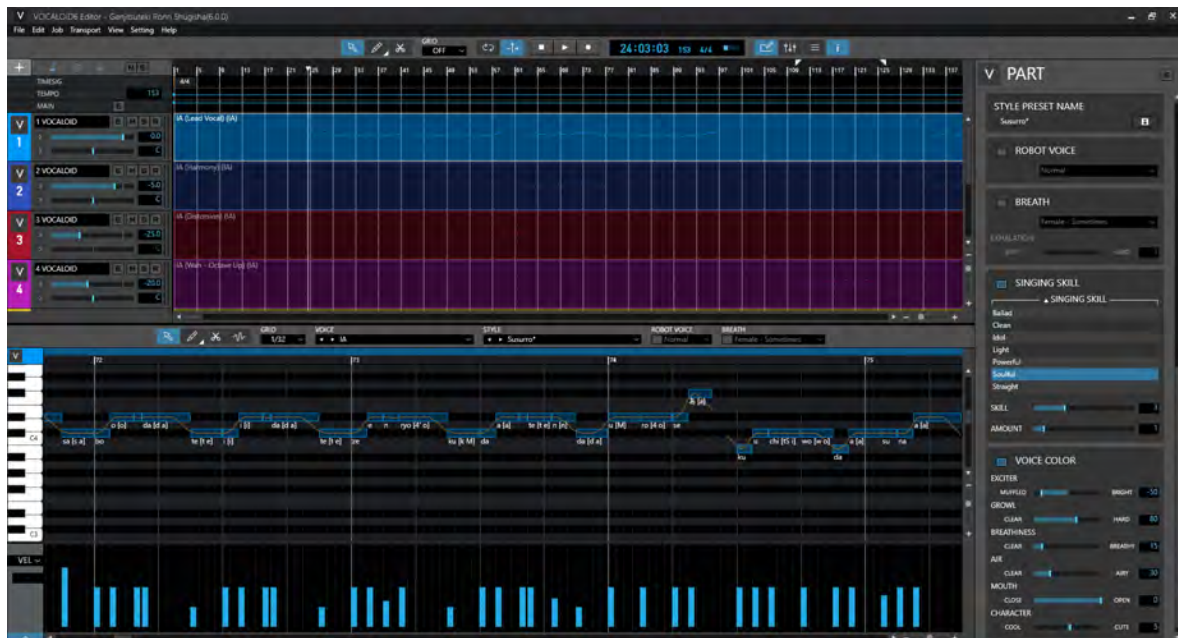
El editor del estilo de canto permite editar desde cero el “estilo” del canto que interpretará la voz en todo el track VOCALOID. En cuanto a los estilos principales que ofrece el *software* se muestran los estilos denominados ROBOT VOICE, BREATH, SINGING SKILL y VOICE COLOR. Este tema musical no contiene un filtro de voz robótica ni tampoco el control de simular respiraciones por frase sintetizada, por lo que no se ha utilizado ni *robot voice* ni *breath*. Sin embargo, se ha utilizado el estilo de canto con la opción SINGING SKILL, ajuste que permite añadir un efecto a la pista MIDI alterando las dinámicas y la expresividad en relación a la secuencia con las opciones que dicen *skill* y *amount*. El apartado de VOICE COLOR modifica el “color” de la voz complementándose con los parámetros que se muestran en el editor de notas sin alterar la pronunciación de los fonemas. Entre estos complementos está el *exciter*, que ajusta el tono de la voz y sus ajustes varían desde el más opaco (-64) hasta el más brillante (+63). El *growl* ajusta el nivel de “gruñido” y sus ajustes van desde el más ligero (0) hasta el más duro (127). El ajuste *breath* permite modificar la cantidad de respiración emulada mezclada con la pronunciación y el tono de las vocales. Los ajustes de esta opción varían desde el más claro (0) hasta el valor más brillante (127). La opción *air* controla la cantidad de expiración mezclada con la pronunciación y sus ajustes varían desde el más ligero (0) hasta el más aireado (127). El ajuste *mouth* controla la dicción, que emula el nivel de apertura de la boca en la pronunciación de las palabras. Los ajustes de este parámetro van desde el más cerrado (-127) hasta el más abierto (0). El parámetro *character* ajusta el “grosor” de la voz en la pronunciación de las palabras. Al aumentar este valor se produce una pronunciación más

femenina mientras que la reducción de este valor produce una pronunciación más masculina.

Estos parámetros aparecen al usuario tal cual se muestra en la Figura 9.

Figura 9

Información de parámetros de voz en la sesión de VOCALOID 6



Nota. Sesión del VOCALOID 6 con la información de la voz que conforma el canal 1 de VOCALOID.

Para esta canción, se han utilizado estos ajustes para cada uno de los 6 tracks que conforman la melodía que hace la voz virtual principal “IA”. El track 1 corresponde a la voz principal o *lead*. En el apartado de *singing skill*, se ha usado un “preset” que el mismo *software* ofrece al usuario que es *soulful*. Esta configuración lo componen dos subparámetros que se mencionaron previamente: *skill* y *amount*. El primer subparámetro se estableció con el valor de 3 mientras que el segundo subparámetro, en 1. A continuación, se ajustó el *voice color* en orden y siempre revisando cómo reacciona el tono vocal en relación a los cambios que se realizaban. Primero se ha revisado la respuesta de la voz con los ajustes del apartado *exciter*, estableciéndose con el valor de -50. Esto significa que la voz tendrá un tono de base no muy “brillante”. En el *growl*, se ha establecido con el valor de 80, otorgándole un mayor ataque en la pronunciación de los fonemas. El ajuste *breathiness* regula el aire que se inhala por la boca al momento de cantar una frase. En este caso se asignó el valor de 15, indicando

un poca presencia de inhalación en el comienzo del fraseo de las estrofas. El ajuste *air* regula el aire que se expulsa por la boca al momento de cantar una frase. En este caso se asignó el valor de 30, indicando poca presencia de aire en el fraseo de las estrofas cantadas. *Mouth* es un ajuste que regula el control del movimiento de la boca al cantar. En este caso, el valor asignado es de 0. Finalmente, el ajuste *character* oscila entre dos subparámetros: *cool* y *cute*. Tiene como valor asignado +10, lo cual indica que la voz tiene una sonoridad ligeramente delgada.

El track 2 corresponde a la voz secundaria o *chorus* en la cual realiza una melodía que complementa a la voz principal manifestándose en las partes del coro de la canción. En el apartado de *singing skill*, se ha usado nuevamente el “preset” que el mismo *software* ofrece: *soulful*. Esta configuración lo componen dos subparámetros que se mencionaron previamente: *skill* y *amount*. El primer subparámetro se estableció con el valor de 3 mientras que el segundo subparámetro en 1. Después se ajustó el *voice color* en orden y siempre revisando cómo reacciona el tono de la voz virtual en relación a los cambios que se van realizando. Se ha revisado la respuesta de la voz con los ajustes del apartado *exciter*, estableciéndose con el valor de -45. Esto significa que la voz tendrá un tono de base no muy “brillante” ya que es una segunda voz con la función de complementarse con la primera voz. Finalmente, en el ajuste del *growl*, se estableció el valor de 80. Para el ajuste *breathiness* se asignó el valor de 20, indicando poca presencia de inhalación en el comienzo del fraseo de las estrofas. El ajuste *air*, *mouth* y *character* mantienen los mismos valores asignados en el track 1.

El track 3 corresponde a la voz que se escucha en el precoro de la canción que se encuentra únicamente antes del primer coro, ubicado aproximadamente desde el minuto 1:30 hasta el minuto 1:37. Un detalle que diferencia esta sección de las anteriores es el uso del paneo como un “call & response”, es decir, la voz virtual entona una frase o verso con el

audio paneado a la derecha y el siguiente verso se encuentra paneado en el lado izquierdo directamente como si se tratase de una respuesta al verso entonado. Los valores que conforman este paneo son de -63 para el lado izquierdo y -64 para el derecho. En el apartado de *singing skill*, se ha usado el mismo “preset” *soulful* como base al igual que los tracks anteriores y valores asignados para el *skill* y *amount*. Después se ajustó el *voice color* ordenadamente. Primero se ha revisado la respuesta de la voz con los ajustes del apartado *exciter*, estableciéndose con el valor de 0. En el *growl*, se ha establecido con el valor de 100, otorgándole un mayor impacto en la pronunciación de los fonemas a nivel general. En el ajuste *breathiness* se asignó el valor de 0, indicando que la presencia de inhalación en el comienzo del fraseo de las estrofas es inexistente. El ajuste *air* tiene como valor asignado la cantidad de 30, indicando un poco más de presencia de aire en el fraseo de las estrofas cantadas a comparación del valor asignado en el track 2. El valor que se le asignó al ajuste de *mouth* es de 0 mientras que el ajuste *character* tiene un valor de +20. Un agregado que tiene este canal es el uso de un panel de efectos, en el cual se puede utilizar los efectos de ganancia o *gain*, *de-esser*, compresión, ecualizador, distorsión o *distortion*, *chorus*, *phaser*, efecto de trémolo, paneo automático o *auto pan*, *delay* y *reverb*. En la Figura 10 se puede observar que todos los efectos que se han utilizado están resaltados de color turquesa. Sin embargo, algunos efectos como la compresión, ecualización y reverberación se tratan mejor en una DAW ya que ofrecen múltiples opciones de edición. El único efecto relevante para replicar la voz en el *software* es uso del efecto de *distortion*, asignándole un valor de +4 en el *drive*. En *lo-fi* y *lo-fi amount* se asignaron los valores de 1 y, finalmente, el *output* tiene un valor de 12 decibeles.

Figura 10

Información de efectos vocales del canal 3



Nota. Sesión del VOCALOID 6 con la información de la voz. Se visualiza el editor de estilo de canto en la parte derecha y el panel de efectos.

El track 4 corresponde a la voz con registro agudo que se escucha durante el puente de la canción, aproximadamente entre el minuto 2:52 hasta el minuto 3:13. Este track no emplea ningún “preset”, por lo que el apartado de *singing skill* se encuentra deshabilitado. Las notas introducidas en el editor se han tenido que reordenar en esta parte de la canción, ya que la voz muestra continuos cambios de articulaciones, melismas y demás ornamentos vocales que han requerido ajustes en la entonación. A continuación, se pasa a editar el color de voz en el apartado de *voice color*. El *exciter* tiene un valor asignado de -20, es decir, un poco más “amortiguado”. El *growl* tiene como valor asignado el 0, mientras que el *breathiness* tiene un valor asignado de 25, lo cual indica que va a contener poca presencia de inhalación. Para el ajuste *air*, se ha designado el valor de 40 indicando un poco más de presencia de aire en el fraseo de los versos. Este ajuste se complementa con el ajuste *mouth*, que se encuentra con el valor de 0. Finalmente, el ajuste *character* tiene el valor asignado de -10, es decir, una voz ligeramente más gruesa. Dentro de lo que abarcan los efectos utilizados en este track, se

conoce que hay un efecto “wah” que se ha aplicado sobre la pista vocal. Sin embargo, el *software* no ofrece dicho efecto de sonido. De la misma forma que el track 3, los efectos se pueden trabajar mejor en una DAW al importar el track exportado en audio con el formato WAV. Para tener un efecto similar dentro del *software*, se ha utilizado un *phaser* de tipo 2 como se muestra en la Figura 11.

El track 5 corresponde a la voz con registro más grave que se escucha durante el puente de la canción que corresponde al mismo intervalo de tiempo mencionado en el track 4. La melodía de la voz virtual en este track está a una octava más abajo que el track anterior y tampoco emplea un “preset” como base, por lo que el apartado de *singing skill* se encuentra desactivado. Las notas introducidas en el editor se han tenido que reordenar exactamente de la misma forma que el track 4. Al editar el color de voz en el apartado de *voice color*, se han ajustado ligeramente los valores del estilo de canto. El *exciter* tiene un valor asignado de -20, es decir, un poco más “amortiguado”. El *growl* tiene como valor asignado 80, mientras que el *breathiness* tiene un valor asignado de 15, lo cual indica que la voz virtual va a tener un timbre poco airoso, pero con impacto en la pronunciación de las palabras. Para el ajuste *air*, se ha designado el valor de 30 indicando un poco más de presencia de aire en el fraseo de los versos si se compara con el track 4. Este ajuste se complementa con el ajuste *mouth*, que se encuentra con el valor de 0. Finalmente, el ajuste *character* tiene el valor asignado de -10, es decir, una voz ligeramente con más cuerpo. Los efectos utilizados en este track es el mismo utilizado en el track 4. Sin embargo, el *software* no ofrece dicho efecto de sonido. De la misma forma que el track 3 y el track 4, los efectos se pueden trabajar mejor en una DAW al importar el track exportado en audio con el formato WAV.

Figura 11

Información de efectos vocales del canal 4



Nota. Se visualiza el editor de estilo de canto en la parte derecha y el panel de efectos.

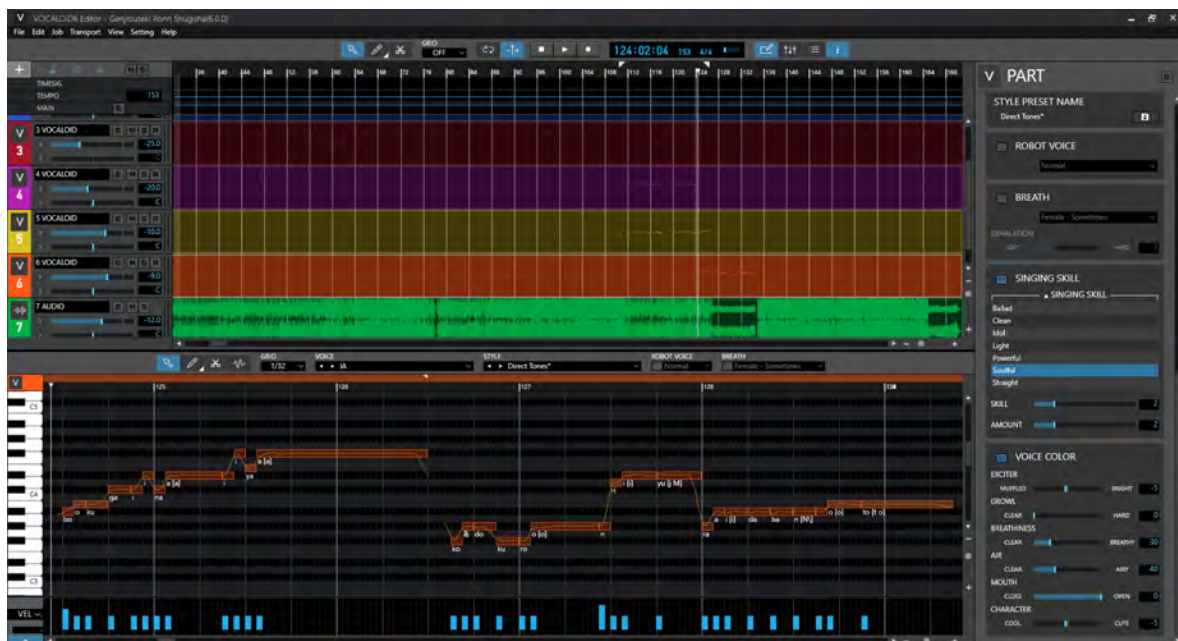
Finalmente, el track 6 corresponde a la sección final del puente de la canción y comienzos del coro en la estructura del tema. Esta parte contiene pequeñas bordaduras y melismas para simular el mismo carácter que el track 4 y 5 para que haya continuidad. Aquí se ha usado un estilo de canto como base que el *software* ofrece: *Direct Tones*. Este “preset” de estilo de canto ofrece una configuración base por defecto que lo conforman los demás subparámetros de la parte vocal, es decir, lo conforman *robot voice*, *breath*, *singing skill* y *voice color*. Esta opción se puede ver encima del editor de notas en donde dice *style*, ubicado a la misma altura que el *grid*. Después se modificó los valores de los ajustes de *soulful*, en el apartado *singing skill*, con los valores de 2 y 2 en *skill* y *amount* respectivamente.

Posteriormente se ajustó el *voice color* en orden. Se ha revisado la respuesta de la voz con los ajustes del apartado *exciter*, estableciéndose con el valor de -5. Esto significa que la voz tendrá un tono de base casi neutral ya que es una voz específica que va a sonar con pocos instrumentos en la que busca generar contraste con el track 4 y 5. Esto también se refleja en el ajuste del *growl*, en el cual se estableció el valor de 0 para transmitir un sonido limpio.

Para el ajuste *breathiness* se asignó el valor de 30, indicando poca presencia de inhalación en el comienzo del fraseo de las estrofas. En el caso del ajuste *air*, se estableció el valor de 40, ya que lo que se busca es que la exhalación simulada por la voz virtual sea parte del contexto sonoro reforzándose por un efecto de *reverb* para un ambiente que envuelva la soledad. El valor asignado para *mouth* nuevamente es de 0 y para *character* es de -5.

Figura 12

Información de efectos vocales del canal 6



Nota. Se visualiza el editor de estilo de canto en la parte derecha.

Los 6 tracks VOCALOID presentan un trato complejo en lo que abarca al estilo de canto. Este tipo de creaciones musicales en las que se utiliza síntesis vocal, se consideran como “música posdigital”. En el artículo “Decuantización, maximalismo y errores forzados. Análisis de los tres códigos dominantes de la música posdigital mediante la extracción de sus principales factores prácticos de codificación” de Carlos Anselmo del Pino, la música posdigital es una forma de música electrónica caracterizada por el exceso de procesos y sonidos digitales como un conjunto de las interacciones con el *software* (2022, p. 118). De esta forma, añade un nuevo valor y apreciación de la música, dejando nuevas formas de experimentación para compositores e incluso cantantes en sí. Se sabe que la música es una

forma de expresión del ser humano, por lo tanto, es considerado un arte por mantener la armonía con sus elementos que lo conforman. En este caso, la función del uso que se da a estos “cantantes virtuales” consiste en crear ese impacto con la audiencia, pero no estrictamente un vínculo con el personaje, sino con el autor y lo que transmite con su obra. Kenmochi menciona que el *software* de síntesis vocal es un nuevo instrumento musical que permite a los compositores transmitir sus emociones de forma más directa al oyente (2012, p. 5388). El uso de estos ‘cantantes virtuales’ no posee límites a una simple creación musical, ya que su objetivo esencial no es la imitación de una voz humana con la capacidad de cantar.

1.2. Análisis de la canción “DOUKEI”

Esta canción fue originalmente compuesta por la banda de rock japonesa BACK-ON cuyo proceso de mezcla y masterización de la pista vocal virtual fue realizado por Out of Service. Este tema fue producido por el motor de VOCALOID 3 y publicado el 5 de noviembre del 2014 en redes como *YouTube* y *Nico Nico Douga* con una duración total de 3 minutos y 39 segundos, siendo parte del álbum “IA/03 –VISION-” (2012) e “IA SUPER BEST –THE ARTISTS- (2022). Esta canción también ha tenido una aparición en el videojuego rítmico “IA/VT –COLORFUL-” (2015) para la consola portátil *Play Station Vita*.

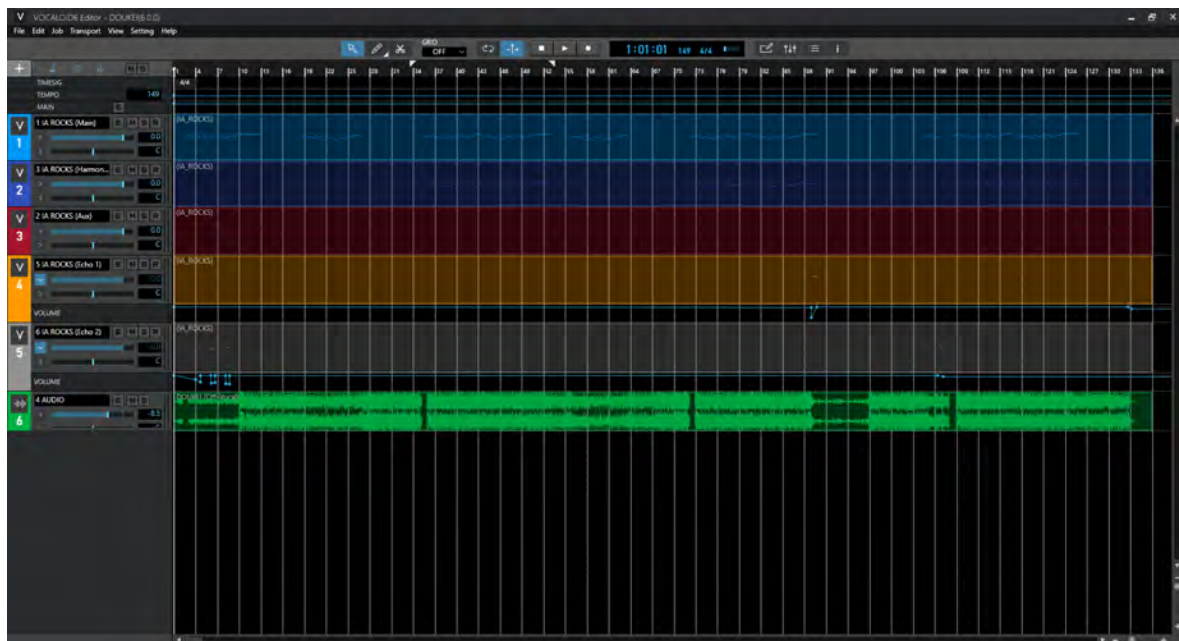
Esta canción pertenece al género *digital rock* o también llamado ‘rock electrónico’, que consiste en la combinación de elementos musicales del rock y la música electrónica con instrumentos que se encuentran típicamente en ambos géneros. El rock electrónico generalmente fusiona elementos de otros estilos musicales, como *punk rock*, *hip hop*, *techno* y el *synth-pop*.

Para llevar a cabo el análisis de la canción a detalle, también se han replicado las secciones vocales en el *software* VOCALOID 6 en su versión *standalone* para visualizar lo que está pasando con el manejo de las voces virtuales, los parámetros que permite el propio *software* para que las voces tengan el tono que se busca y los efectos de sonido que se les

puede aplicar para contextualizar mejor el carácter que se le da a la canción. En la Figura 13 se puede observar la información del cantante virtual utilizado, el estilo de canto del *voicebank* y los canales que conforman la sesión siendo 6 canales en total: 5 canales VOCALOID con su respectivo editor de notas y 1 canal de audio.

Figura 13

Interfaz general del usuario en VOCALOID 6



Nota. Sesión del VOCALOID 6 de la segunda canción sin el menú lateral derecho

La canción tiene como único banco de voz a “IA ROCKS”, también creada por 1st PLACE Co., Ltd. Este *voicebank* es una variante de la voz original de IA y tiene la misma característica de tener un parecido con la proveedora de voz Lia, salvo que tiene la diferencia de ser un *voicebank* orientado para canciones que requieran un tono más agresivo. Esta voz virtual también presenta un comportamiento un poco diferente en comparación con el *voicebank* IA y eso se verá reflejado en el renderizado de tono. De igual manera, simular los adornos vocales en el editor no es suficiente con solo insertar las notas, sino también realizar ligeras modificaciones a la línea melódica para que la voz responda poco a poco hasta que se obtenga el sonido que uno espera.

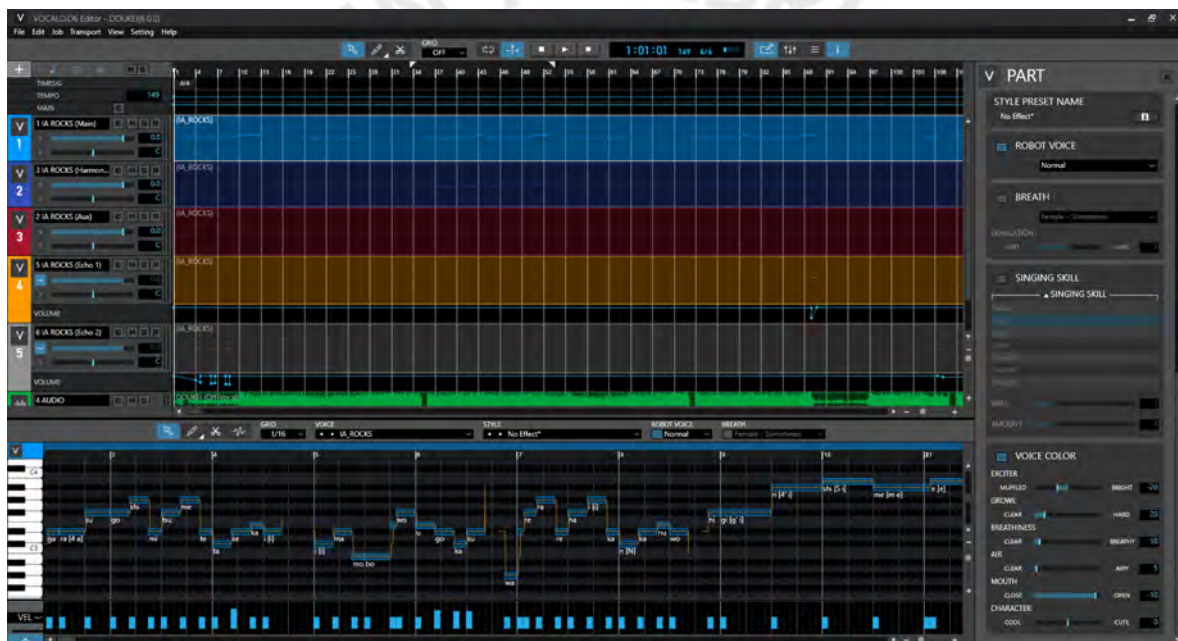
Para esta canción, se han utilizado estos ajustes para cada uno de los 6 tracks que conforman la melodía que hace la voz virtual principal “IA ROCKS”. El track 1 corresponde a la voz principal o *lead*. El apartado de *singing skill* y *breath* no se han utilizado para este track, siendo los únicos ajustes activos las opciones *robot voice* y *voice color*. El primer ajuste altera la renderización de las notas en el editor convirtiendo los ataques para que sean siempre directos. En la Figura 14 se puede observar el comportamiento de la entonación con el ajuste *robot voice* y, a diferencia de la otra canción, la entonación no posee curvaturas al conectarse con las demás notas como si no existiese un *bend* o *glissando* como por ejemplo en el compás 38 del editor que puede visualizarse que, entre nota y nota, la línea amarilla de entonación no es en líneas curvas como en el anterior tema, sino es con menor *legato*. Esta configuración del comportamiento vocal permite simular el efecto sonoro de voz robótica que se dividen en *soft*, *normal* y *hard*. Para este caso, se ajustó en *normal* para que encaje con el estilo musical y no suene demasiado robótico, pero tampoco sin sonar muy “limpio”. Después se ajustó el siguiente parámetro activo: *voice color*. En este apartado, el *exciter* está asignado con el valor de -20, es decir, el timbre de la voz emulada va a tener ligeros toques de “opacidad”. En los ajustes del *growl* se asignó el valor de 20, indicando que la voz virtual no emulará mucho efecto de “gruñido”, manteniendo una pronunciación más limpia sin frases levemente marcadas con este ajuste. A la opción *breathiness* se asignó el valor de 10, indicando que antes de cada frase, el *software* simulará sutilmente la respiración. Para el ajuste *air* se asignó el valor 5 para que el motor de síntesis emule sutilmente la exhalación al concluir una frase melódica. En la opción que dice *mouth*, el valor asignado es -10, ya que, al tener el efecto de *robot voice* activado, ocasiona una mayor presencia en las vocales abiertas al entonar las notas. Finalmente, el ajuste *character* se mantiene con el valor de 0.

El track 2 corresponde a la voz secundaria o *chorus* en la cual realiza una melodía que complementa a la voz principal del track 1 solamente en las partes del coro de la canción.

Igual que el track 1, esta parte vocal también tiene el efecto de *robot voice* establecido en *normal*. En el apartado de *breathiness* y *singing skill* no se han usado un “preset”, ya que estas funciones se han mantenido desactivadas. Después se ajustó el *voice color* manteniendo casi los mismos valores asignados en el track 1, ya que la voz solo cumple la función de “refuerzo armónico” para complementar la primera voz. El único valor diferente es el ajuste del *air*. En este track, el valor asignado es de 10, enfatizando un poco más la presencia de aire en la pronunciación.

Figura 14

Información de parámetros vocales en VOCALOID 6



Nota. Sesión del VOCALOID 6 con la información de la voz que conforma el canal 1 de VOCALOID.

El track 3 corresponde a una melodía que realiza la cantante virtual en tres momentos específicos: después de la intro del tema, después del primer coro antes del puente y después de la repetición del coro tras cambiar la tonalidad. En este track se ha utilizado un ajuste más personalizado en las propiedades de cada nota insertada en el editor. En la Figura 15 se puede visualizar un menú que se sitúa en la parte lateral derecha de la imagen con el nombre de *note inspector* o también llamado “inspector de notas”. Al seleccionar una nota que se insertó en el editor, al hacer clic en el botón “inspector” en el área superior derecha de la pantalla de la

ventana principal se muestra el panel *note inspector* en el lado derecho de la ventana principal. *Note inspector* puede mostrar información y cambiar la configuración de la nota seleccionada. Para este caso, se ha editado el tono de la pronunciación de la nota seleccionada en la cual se escogió la opción *clear* en el caso del ataque ya que se busca emular la gesticulación de la boca para la pronunciación. Este ajuste con el *note inspector* se ha realizado con cada una de las notas del track. Además, el track contiene el efecto *robot voice* con la configuración activada en *normal* y posee cierta similitud con respecto a los otros track anteriores. En los ajustes del *voice color*, el valor del *exciter* se estableció en -20, siendo el mismo valor del track 1 y del track 2. En el caso del ajuste del *growl*, el valor es de 0 debido a que se entona una melodía con una vocal abierta con el fonema /o/. Una entonación sin gruñido da la claridad necesaria para que se complemente con la parte instrumental de fondo. El ajuste *breathiness* se mantiene con el mismo valor que los tracks anteriores con el valor de 10. El ajuste de *air* ha aumentado en comparación con los demás tracks mencionados siendo el valor de este track el de 25. Se desea una textura que no sea con el mismo estilo que la voz principal ni la voz secundaria, por lo que se ha decidido que el estilo para la entonación de este track sea más “airosa” al concluir las notas con mayor duración y con un mayor intervalo de silencio antes de comenzar otra frase. Los ajustes de *mouth* para este track se mantiene en 0, al igual que el ajuste *character*. A pesar de poder encajar dentro del track 2, se ha decidido trabajarse en un canal diferente por un tema de efectos de sonido que se han utilizado. El más relevante corresponde al uso del *delay*, en el cual el tiempo asignado de retraso es mayor a comparación del tiempo que se utiliza en los tracks 1 y 2.

Figura 15

Información de efectos vocales en el canal 3 de VOCALOID 6

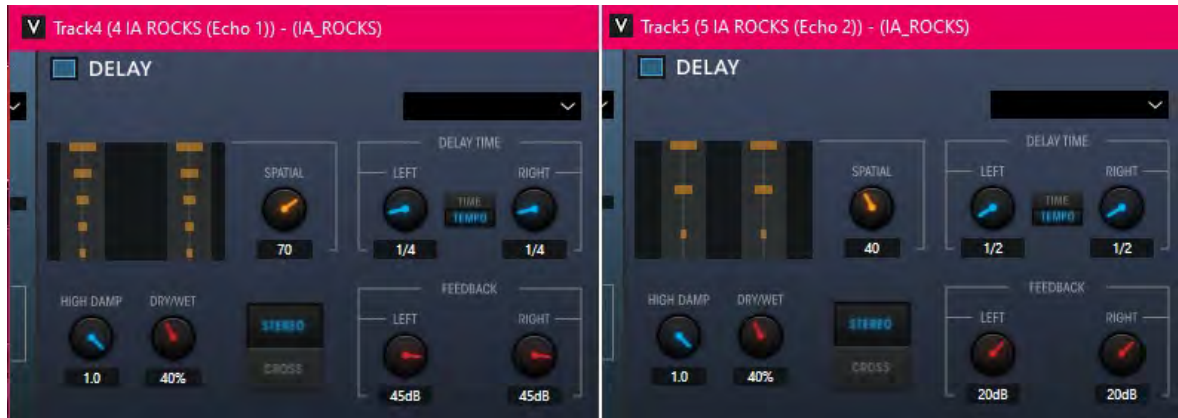


Nota. Se visualiza el editor de estilo de canto en la parte derecha y el panel de efectos.

El track 4 y 5 corresponden a canales secundarios para reforzar los tracks principales con algún efecto de sonido aplicado a alguna sección en concreto. En este caso, el track 4 solamente contiene la última sílaba de la nota ubicada antes de terminar el segundo coro para pasar al puente. Dentro de los parámetros del estilo de canto, solamente tiene la opción *robot voice* activada. Lo que quiere rescatarse de este track es el efecto de *delay* y es el motivo por el cual este track solo destaca la última sílaba en un momento específico. Además, el track 4 utiliza la automatización del volumen. Su función es la misma como si se utilizase otra DAW y en este caso, se busca que sea la repetición del sonido causado por el *delay* lo que resuena de fondo. En el caso del track 5, además de utilizar el *robot voice*, si tiene los mismos parámetros en el *voice color* que el track 1. Además, utiliza el efecto de sonido del *delay* pero se diferencia del track 4 en el intervalo de tiempo.

Figura 16

Configuración del efecto de delay de los tracks 4 y 5



Replicar el comportamiento vocal de esta canción permite tener una referencia visual del proceso que se tuvo durante la composición. El compositor debe saber lo que está buscando transmitir con la pieza y debe ser congruente con el concepto de la canción. Este tema tiene elementos del *rock*, *techno* y secciones de rap, cantadas por un cantante real durante el precoro. En el coro, la voz virtual con el efecto *robot voice* es congruente con la temática y género de la canción, por lo que el cambio causa un contraste en la audiencia y le da mayor dinamismo en el clímax situado al finalizar el coro.

Capítulo 2. Impacto de VOCALOID en 20 jóvenes limeños entre los 18 y 48 años

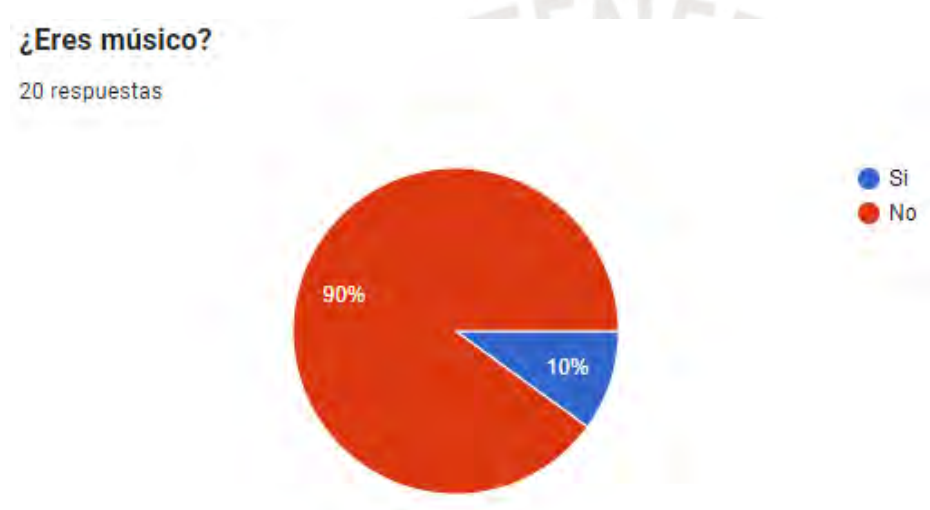
En el mundo existen numerosas composiciones creadas utilizando VOCALOID, al igual que otros programas similares con la misma finalidad: experimentar con diferentes registros y estilos musicales empleando de diferentes formas las voces virtuales que se les ofrece. En el capítulo anterior se analizaron dos canciones y se pudo comprobar que una única voz puede tener distintos tonos con los ajustes que puede hacerse en el *software* según el gusto del usuario. Sin embargo, ¿qué impacto tiene este estilo de canciones en los oyentes? La música es un producto que las personas consumen a diario. Si bien es cierto, los dos temas mencionados del capítulo anterior están en el idioma japonés, siendo Japón el país en donde se popularizó este *software* en un inicio. La popularización de VOCALOID está vinculada con la toma de decisiones que se ha tomado para llevar a cabo la ilustración de los cantantes virtuales relacionada a la “cultura del anime” y del entretenimiento. Esto se ve reflejado en la gran mayoría de los videos promocionales o *videoclips* que acompañan la canción, siendo el principal atractivo para las personas que ya están en contacto con este tipo de ilustraciones. El impacto en la audiencia se da a través del tipo de interacción que se tiene con el *software* presentado principalmente como un “*software* para simular el canto humano” pero también visto como “un medio de entretenimiento para conectar con la audiencia”.

Para llegar a estas afirmaciones, se ha tenido que realizar una encuesta para determinar la forma en la que las canciones con *software* que permite el uso de cantantes virtuales ha tenido un impacto en las personas. Se tomó una población de 20 personas limeñas dentro del margen de 18 y 48 años de edad para realizar el estudio que refleje cómo ha sido la recepción por parte de los oyentes, determinando si hay una respuesta positiva o negativa sobre el uso de cantantes virtuales. A los encuestados se les ha presentado las canciones “A Realistic Logical Ideologist” y “DOUKEI”, las mismas canciones que han sido analizadas en el Capítulo 1. Entre las preguntas generales, se les preguntó si han escuchado con

anterioridad sobre el uso de estos *softwares* para la emulación del canto y si el realismo en las voces virtuales es un elemento importante y necesario. Al hablar del realismo en la canciones, se refiere a que se consiga transmitir con el mismo sentimiento lo que una voz humana puede interpretar con cada palabra. Además, para la mayoría de personas, VOCALOID es el *software* más utilizado y conocido entre los encuestados, además que el realismo es un elemento fundamental que va a determinar los gustos y las preferencias de cada uno.

Figura 17

Gráfico circular que presenta a los músicos y no-músicos



Es importante conocer si las personas encuestadas son músicos o no. El músico es la persona que, además de saber tocar y dominar su instrumento, conoce lo que está sucediendo en la interpretación, composición, arreglo y producción musical de una forma más analítica. Sin embargo, la perspectiva con la que se deberían ver estas piezas en esta sección ahora será como un “producto” que vende, es decir, las personas oyentes son las que deciden porque son los que reciben el producto musical. Si el producto es de su agrado, van a continuar escuchando las canciones y es un patrón que se suele dar. La diferencia está en cómo se maneja la voz en este tipo de música. Hay registros que para una voz real sería muy complicado pero que al mismo tiempo no es cuestionado o criticado, sino aceptado como el producto que es. En esta encuesta, del total de 20 personas, 18 personas no son músicos

mientras que 2 personas si lo son, evidenciando que la mayoría de personas que son oyentes no se dedican completamente a la música sin cuestionar el proceso que conlleva el producto musical.

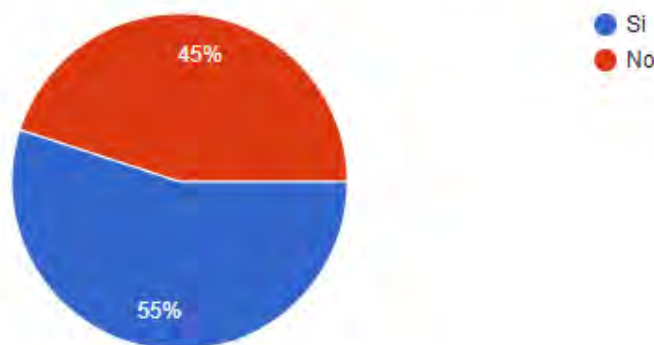
Se sabe que con la llegada de VOCALOID 2, la ilustración del *voicebank* de Hatsune Miku se convirtió en la nueva forma de vender las voces virtuales debido a que las personas asocian la voz virtual con la imagen representativa del cantante y esto establece un vínculo entre la ilustración y las personas que consumen el producto musical. Una parte de este “producto” también incluyen los videos promocionales de las canciones. Los *videoclips*, al igual que las canciones, eran creadas por las personas para el *voicebank* que protagonizaba la canción hasta el punto de inventarle una biografía y darle una personalidad propia. Las redes permiten romper barreras sociales en lo que abarca la creación musical fomentando la participación y la colaboración de los usuarios desde la creación musical, la composición de la letra, el *videoclip*, la imagen y arte conceptual e historia o temática que llevaría la canción o album. Es probable que la popularidad o reconocimiento que tuvo VOCALOID en sus inicios no haya tenido el mismo impacto si se hubiese tomado otro rumbo para publicitar las voces virtuales.

Figura 18

Gráfico circular sobre la familiaridad de entrevistados con cantantes virtuales

¿Has escuchado/leído sobre la síntesis vocal o "cantantes virtuales"?

20 respuestas



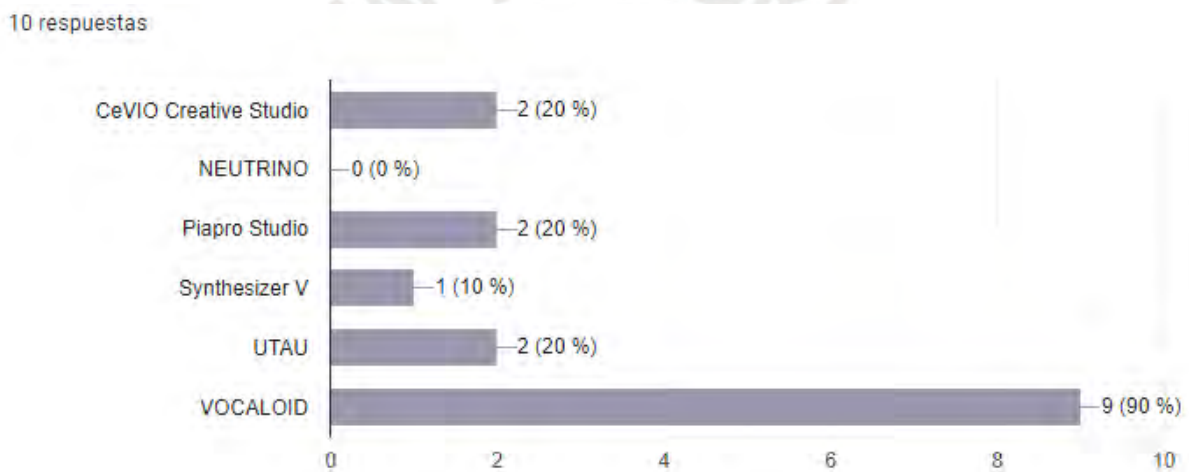
Para hablar del “producto musical” se les ha pedido a los usuarios que, para los que han escuchado sobre el uso de “cantantes virtuales”, señalen dentro de la encuesta los *softwares* relacionados a la síntesis vocal que les sean conocidos de la lista en la Figura 19. De la población de 20 personas, solamente 10 personas conocen algunos programas para síntesis vocal mencionados en la lista. Entre las opciones está CeVIO Creative Studio, que al igual que VOCALOID, es un *software* para síntesis vocal enfocada en el canto pero también en emular el habla. Solo 2 personas de un total de 10 han marcado esta opción. A diferencia de VOCALOID, la interfaz del usuario solo se encuentra en el idioma japonés. La siguiente opción en la lista es NEUTRINO, el menos conocido por los usuarios ya que ninguno de los encuestados ha marcado este *software*. Este *software* es un sintetizador de voz creada para el canto pero que utiliza una “red neuronal⁴” como motor de síntesis basado en una inteligencia artificial. En el motor de síntesis estima el tiempo, tono y la calidad de la voz de la partitura. Al estar basado en redes neuronales, no posee una interfaz de usuario como los demás *softwares* pero permite interpretar archivos en formato XML. El programa que continúa en la lista es Piapro Studio, *software* para síntesis vocal desarrollado por Crypton Future Media con la característica que se enfoca para su uso dentro de una DAW. Originalmente utilizó VOCALOID como su motor de síntesis. Posteriormente, Piapro Studio tiene su versión *standalone* con el que busca “solucionar” las limitaciones que ofrece el *software* original VOCALOID. Igual que en el caso de CeVIO Creative Studio, solamente 2 personas de un total de 10 han marcado esta opción.. El siguiente *software* en la lista es Synthesizer V creado por Dreamtonics, Co. Ltd. Utiliza un motor de síntesis híbrido que utiliza inteligencia artificial para crear voces similares a las de los cantantes reales. Para este *software* solo ha habido 1 persona que ha marcado esta opción. El *software* que le continúa en la lista es UTAU, sintetizador de voz japonesa desarrollada por AMEYA/AYAME. El nombre UTAU

⁴ Modelo de computación que simula el funcionamiento del cerebro humano, permitiendo a las máquinas aprender y tomar decisiones de manera similar a como lo hace el ser humano.

significa Singing Voice Synthesis Tool. Este *software* está inspirado en VOCALOID con el agregado que cualquier persona puede utilizar sus propias muestras de voz o desonido en el programa para múltiples propósitos que generalmente son relacionados al canto. Solo 2 personas han señalado que conocen este *software*. Finalmente, el último *software* en la lista es VOCALOID que abarca a la mayoría de la población seleccionada evidenciando que es el más popular entre los usuarios y oyentes contando con 9 personas que han marcado este programa.

Figura 19

Gráfico de barras sobre la variedad de motores de síntesis vocal que hay



Como se pudo ver, VOCALOID no es el único *software* que se dedica a trabajar con cantantes virtuales, sino existen muchos más con la misma función y ofreciendo alguna propiedad extra que le conceda un lugar en el mercado. Sin embargo, su objetivo consiste en el mismo: emular el canto humano. El canto humano está lleno de dinámicas y expresiones que transmiten un sentimiento. El realismo en las voces es un elemento importante porque se sabe que todas las canciones creadas por cualquier sintetizador de voz no utilizan una voz real. Si tomamos la tesitura de una voz real como significado de la sonoridad realista en su estado más puro, entonces la síntesis de voz no se va a igualar de ninguna forma a una voz real, es decir, bajo esta premisa no es posible que una máquina pueda competir o llegar a igualarse en su totalidad con una característica que es propiamente biológica, parte del propio

ser humano. En la Figura 19 se puede ver el porcentaje de las personas que consideran que el realismo en la voz es importante, siendo 8 personas los que consideran que el realismo en la voz no es importante mientras que para las otras 12 personas sí es un elemento importante a tener en cuenta.

Figura 20

Gráfico circular sobre la relevancia del realismo vocal



2.1. Análisis de las respuestas sobre cada una de las preguntas

Las preguntas que se plantearon proponen demostrar el impacto y respuestas de las personas sobre cada uno de los elementos que conforman el producto musical. Por eso, a cada encuestado se le ha realizado 6 preguntas en relación a su percepción sobre las canciones “A Realistic Logical Ideologist” y “DOUKEI”. Para referirse a estas dos canciones se les llamará “canciones de muestra” ya que son los temas musicales que se les ha hecho escuchar a los encuestados. A la canción “A Realistic Logical Ideologist” se le llamará “canción #1” por ser el primer tema musical que se les presentó a los encuestados. Finalmente, para la canción “DOUKEI” se le va a referir “canción #2”, ya que ha sido la segunda canción que los encuestados han tenido que escuchar.

La primera pregunta que se ha compartido con los encuestados ha sido: “¿Cuál de las dos canciones ha sido más de tu agrado?”. Del total de la población que abarca 20 personas,

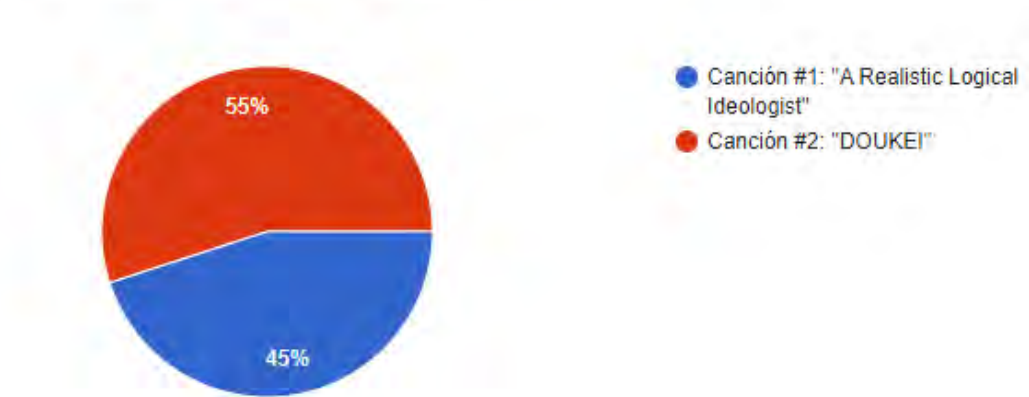
11 personas han preferido la canción #2 mientras que 9 personas, la canción #1. Cabe mencionar que las canciones han sido presentadas con su respectivo video promocional con el que se publicó originalmente. La canción #1 posee una mayor suavidad en el tono de la voz en comparación con la canción #2 que tiene un carácter más enérgico en las voces. Sin embargo, la canción #2 tiene un precoro en el que participa un cantante real, ofreciendo un aire más fresco y original a la canción.

Figura 21

Gráfico circular sobre la preferencia hacia la canción #2

¿Cuál de las dos canciones ha sido más de tu agrado?

20 respuestas



Como se ha mencionado antes, cada canción no ha sido publicada independientemente como una canción sola, sino como un “producto musical”, en el cual, además de la música, también refleja con la misma jerarquía el video promocional que acompaña a las canciones. Muchos de los *videoclips* que circulan en el internet permiten visualizar la narración de lo que se narra en la letra y así tener una explicación más clara tomando en cuenta que las canciones de muestra están en el idioma japonés. Además, muchos de estos videos son animados por los mismos usuarios y otros, por algún artista afiliado con algún sello discográfico. En este caso, la canción #1 ha sido animada por una persona que grafica lo que la canción relata para darle un sentido. En la canción #2, el *videoclip* fue realizado por un ilustrador que pertenece a lo que se conoce como “IA PROJECT”,

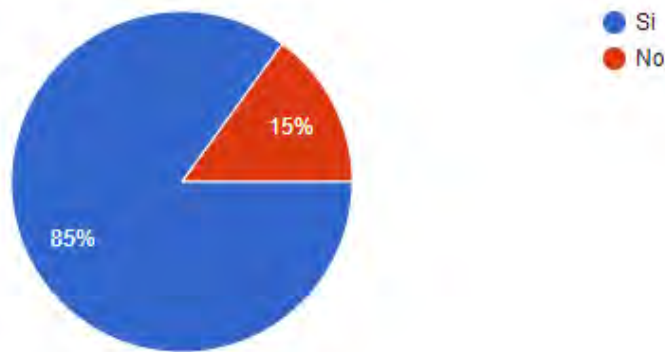
organización de 1st PLACE Co., Ltd. encargada de continuar promoviendo la actividad musical y comercial a través de eventos, *merchandising* o desde su misma página web relacionada al uso de su *voicebank* IA. El *videoclip* de la canción #2 no tiene el mismo estilo artístico, pero encaja con la intención de lo que dice la letra y lo que transmite la música. Por eso la segunda pregunta es: “¿Consideras que el videoclip que acompaña a la canción es importante para llegar a la audiencia?”. De las 20 personas encuestadas, 17 personas consideran que el video promocional sí es importante para conectar mejor con la audiencia mientras que 3 personas afirman que un *videoclip* no influye en la forma en cómo van a percibir el producto.

Figura 22

Gráfico circular sobre la relevancia de los videoclips

¿Consideras que el videoclip que acompaña a la canción es importante para llegar a la audiencia?

20 respuestas



La tercera pregunta de la encuesta es: “¿Qué canción consideras que suene más ‘realista’?”. Previamente, se mencionó sobre la importancia del realismo en el uso de las voces virtuales y cómo es que esta característica es importante para la mayoría de los encuestados. A diferencia del uso de instrumentos virtuales, *samplers* y efectos de sonido que se pueden aplicar a una toma, el manejo de las voces virtuales tienden a evidenciar más el margen de error en la pronunciación de algunas frases si no se renderiza correctamente la entonación. Es por eso que existen personas que se dedican específicamente a ser productores

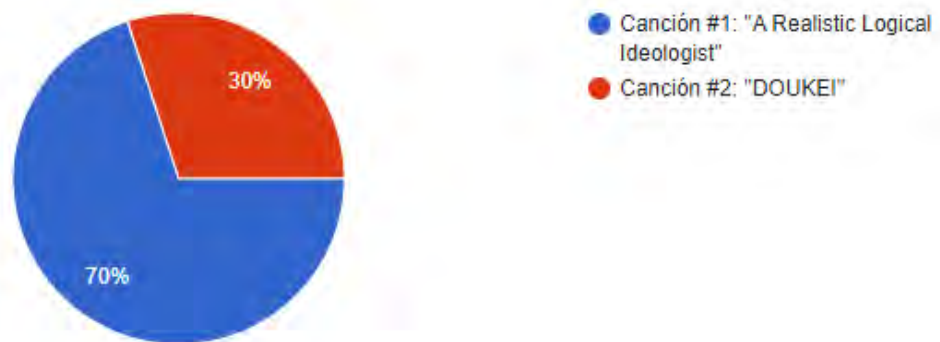
VOCALOID con la tarea de pulir el renderizado de las voces para buscar la mejor sonoridad, pronunciación, distribución de las notas para añadir eficazmente ornamentos y la configuración de los parámetros de la voz para que suene lo más fluído y similar a las cualidades del canto humano. En la Figura 21 se puede ver que del total de los encuestados, 14 personas consideran que la canción #1 tiene una sonoridad más realista que la canción #2, conformado por 6 personas, en cuanto al uso dado en la voz virtual.

Figura 23

Gráfico circular sobre el realismo hacia la canción #1

¿Qué canción consideras que suene más "realista"?

20 respuestas



La cuarta pregunta que se ha preguntado a los encuestados es: “¿Considerarías escuchar música de este estilo más frecuentemente?”. Al mencionar “estilo” uno se refiere a la propuesta musical en cuanto a la decisión de utilizar un cantante virtual en vez que un humano lo cante. Ciertamente un recurso cuyo origen es emular una habilidad humana no va a llegar a parecerse en su totalidad. Hasta la actualidad la mayoría de *voicebanks* son de habla japonesa pero también existen *voicebanks* de habla inglesa y habla hispana. Con el último lanzamiento recientes del *software* VOCALOID 6 y el uso de la inteligencia artificial en su motor de síntesis, es posible que los cantantes virtuales mantengan su timbre vocal y pasen a ser “multilinguaje”. Esto quiere decir que podrán utilizarse en cualquier idioma editando el fonema que se desea que la voz virtual entone ofreciendo nuevas texturas en el canto para

encontrar un estilo propio sin buscar parecerse al canto humano sino encontrar su lugar en el mercado para poder valerse por si mismo. En la Figura 22, se observan las respuestas de los encuestados sobre esta cuarta pregunta. Del total de 20 personas, 14 personas están dispuestas a escuchar más canciones de este estilo con cantantes virtuales, mientras que las otras 6 personas no desean escuchar este estilo de canciones.

Figura 24

Gráfico circular sobre la preferencia del estilo musical vocal



La quinta pregunta de la encuesta es: “¿Consideras que el uso de ‘cantantes virtuales’ puede ser un reemplazo de los ‘cantantes reales’ a largo plazo?”. Esta pregunta que se ha realizado está conectada a la pregunta anterior porque el tener que asumir algo cuyo origen y desarrollo se da un campo digital que pueda superar una característica de origen biológico propiamente humana, no va a poder ser idéntica a una voz real. Un ejemplo de esta idea se ha dado cuando Yamaha comenzó a promocionar el *software*, empleando una etiqueta en la cual intenta sugerir una idea al consumidor: “Human beings are still singing? We are in the digital age! ... It’s time for you to join us!”⁵(2016). El uso de voces virtuales se comporta más como un complemento para transmitir una emoción o alguna idea musical en específico. En la canción #2 se puede ver eso al intercalar en el precoro con un cantante real generando

⁵ La traducción de este eslogan es: “¿Los humanos aún están cantando? ¡Estamos en la era digital! ... ¡Es hora de que te unas!

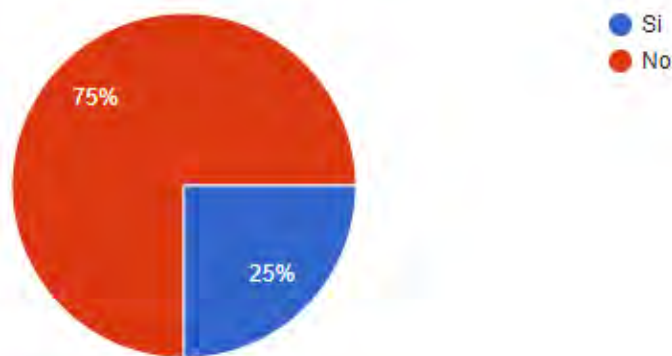
contraste en la textura que abarca la sección musical. En la Figura 23 se observa que, del total de 20 personas encuestadas, 15 personas niegan que los cantantes virtuales pueden ser un reemplazo a largo plazo de los cantantes reales, mientras que las otras 5 personas afirman que si pueden ser un reemplazo a futuro de los cantantes reales.

Figura 25

Gráfico circular sobre el reemplazo de cantantes virtuales sobre los reales

¿Consideras que el uso de "cantantes virtuales" puede ser un reemplazo de los "cantantes reales" a largo plazo?

20 respuestas



Finalmente, la sexta y última pregunta que se ha realizado a los encuestados es: “Crees que el uso de un/una cantante virtual ayuda en la composición de canciones?”. Cabe resaltar que casi la totalidad de encuestados no son músicos, por lo que las respuestas sostienen la postura de la pregunta anterior. Si bien es cierto que los cantantes virtuales se presentaron como “los cantantes del futuro”, funcionan más como una herramienta para la composición musical para crear un estilo propio con su propia estética. Esto le permite posicionarse como “producto” en el mercado. Además, reconocer que es un recurso adicional para la composición de canciones no solo brinda más posibilidades a los compositores, sino que las personas que conforman la audiencia reconocen que no llega a igualar a un ser humano, sino a ofrecer mayor variedad como “producto musical” y como un “recurso para la composición”. Con respecto a la pregunta realizada a los encuestados, en la Figura 24 se puede observar que, del total de 20 personas, 15 personas afirman que el uso de cantantes

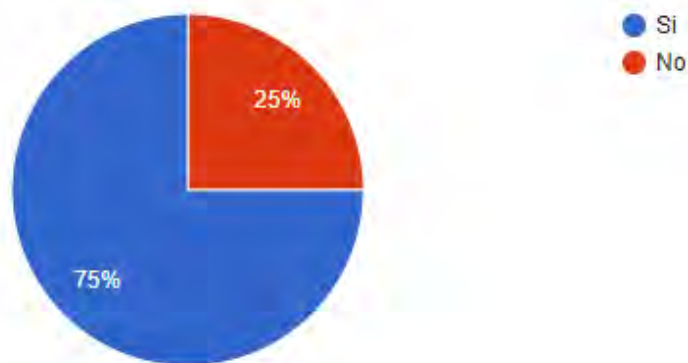
virtuales sí ayuda y aporta en la composición de canciones mientras que las otras 5 personas consideran lo contrario.

Figura 26

Gráfico circular sobre la relevancia en el campo del songwriting

¿Crees que el uso de un/una cantante virtual ayuda en la composición de canciones?

20 respuestas



2.2. Análisis de resultados globales

Las seis preguntas sobre la percepción que causaron las canciones creadas con VOCALOID sí ha tenido un impacto sobre la cantidad de encuestados, permitiendo tener una percepción más clara de la intención que se debe tener con el uso del *software* en la creación musical y preferencias sonoras, considerando lo que el público espera consumir como “producto musical”. La estética que se adopte al momento de producir la canción va a tener un impacto mayor o menor según las preferencias que tenga la audiencia a la que va dirigido. Las primeras canciones con el uso de cantantes virtuales no tenían una imagen que caracterice a la voz virtual y, por lo tanto, no han tenido el mismo impacto que en la actualidad, que es como se conoce el *software*. A partir de VOCALOID 2 se venden los *voicebanks* junto a un avatar en la portada que se asigna a la voz virtual. Para realizar los *videoclips* que se reproducen en conjunto con la canción creada, se utiliza el mismo avatar para conceptualizarlo en la elaboración del video promocional. El tipo de ilustración que se

utilizó toma como base el estilo de animación japonesa denominada “anime” y es parte propia de la cultura del entretenimiento de su país que ha ido popularizándose en todo el mundo. Tener esto en cuenta es importante porque a partir de esta premisa las personas encuestadas han confirmado a través de las respuestas que han dado.

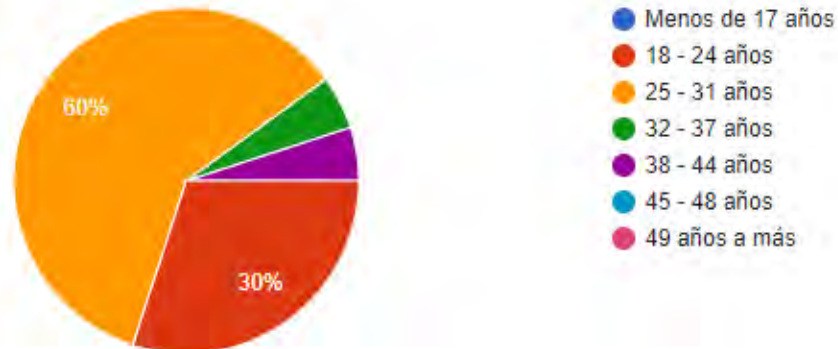
Entre ambas canciones que se han sido presentadas a las 20 personas encuestadas que conforman el 100%, 11 personas corresponden al 55% que tienen mayor preferencia por el estilo que adopta la canción #2. Como tema musical, la canción presenta elementos de la música *techno* tanto en la instrumentación como en su progresión armónica. Como “producto”, la música está vinculada con su respectivo video promocional en el cual una toma la estética característica del anime, mientras que la otra canción tiene un *videoclip* que no se enfoca en el personaje, sino en lo que se busca transmitir con respecto a la canción. Es decir, el *videoclip* de la canción #1 actúa como una extensión de la canción con la función de contextualizar un tema mientras que en el *videoclip* de la canción #2, actúa más como un complemento de la canción. En la Figura 27, se puede ver el margen de edades de las personas encuestadas la mayoría pertenece al rango de 25 años hasta 31 años conformado por 12 personas, 6 personas están dentro del grupo de personas de 18 a 24 años y 1 persona dentro del rango de 32 a 37 años al igual que los de 38 a 44 años.

Figura 27

Gráfico circular sobre el rango de edades de los encuestados

¿Cuál es tu rango de edad?

20 respuestas



La mayoría de las personas encuestadas afirman que para ellos es importante el realismo en las propiedades sonoras de una canción. Se sabe que para la mayoría de las personas encuestadas, la canción #1 tiene una sonoridad más realista en el manejo de la voz virtual. Sin embargo, a pesar de eso, la mayoría de encuestados tiene una mayor afinidad con la canción #2, en la cual se posee un manejo de la voz virtual con una característica sonora más artificial por el efecto *robot voice*. Como ya se ha mencionado antes, la canción #2 tiene en el precoro una sección de rap con una voz real. Si bien es cierto que cambiar de una voz virtual a una real genera contraste en la canción, el uso de una voz virtual se complementa con mayor versatilidad que en la canción #1. Esto ocurre debido a que en la canción #2 el género musical del todo el tema es el *techno*, derivado de la música electrónica en la que se suele utilizar secuenciadores, sintetizadores, *samplers*, teclados y *vocoders*⁶. Debido que su instrumentación tiene como base el uso de instrumentos y herramientas digitales, no se disocia de utilizar una herramienta con una finalidad sonora que también es digital. Es decir, el uso de las voces virtuales se complementa mejor con ciertos tipos de género siendo la música *techno* uno de ellos. Además, la estética de la música *techno*, debido a su

⁶ Analizador y sintetizador de voz considerado como un instrumento musical al tener la capacidad de conectar guitarras y teclados para simular la síntesis vocal.

instrumentación, es futurista o *cyberpunk*. Esta estética está propiamente manifestada en el video promocional de la canción #2 en la que se utiliza un fondo oscuro, iluminación de colores entre ratos del video y el estilo del *glitch art* para las transiciones de video.

En el caso de la canción #1, el *videoclip* adopta la estética del “anime”. A pesar de que la canción tenga un tono vocal más “realista” en contraste con la canción #2, la canción #1 proviene de una combinación de estilos derivados del *techno* y del género del *rock*. Evidentemente, posee más elementos del *rock* y es por eso que la inserción de una voz virtual en un género cuyo origen no es puramente digital, no va a generar una sonoridad tan congruente como una voz virtual que se complementa con su estética. Es decir, el realismo en la voz está presente pero la diferencia está en que en la canción #1 busca parecerse a una voz real dentro del estilo musical que se propone, mientras que en la canción #2 busca que la voz virtual se complemente con la estética musical propuesta. Es por eso que los encuestados tienen mayor preferencia por la canción #2 que por la canción #1 a pesar que la canción #1 tenga una característica vocal más realista.

Entre los acontecimientos que abarcan la historia del *software* VOCALOID, ha habido una voz virtual que se ha mencionado previamente que es la del *voicebank* Hatsune Miku, que significa “el sonido del futuro”. Hatsune Miku ha sido un *voicebank* muy relevante dentro de la evolución de VOCALOID, ya que todo lo demás posterior al lanzamiento de este *voicebank* va a marcar un estereotipo de lo que es VOCALOID tanto a los que utilizan el *software* como para los oyentes. La imagen que se le dio a este banco de voz fue elaborada por Kei Garo, ilustrador de manga⁷ muy comercial en oriente. Según Cuevas, estos elementos que se emplearon tanto en el diseño y elaboración del personaje como en la creación del banco de voz fueron vitales para asegurar el éxito del *software* en sí a nivel mundial, iniciando por Japón (2013, p. 52). La elaboración del diseño del personaje, según la compañía

⁷ Historieta de origen japonés.

Crypton Future Media que durante un tiempo trabajó junto a Yamaha Corporation, fue solo para darle una imagen propia que caracterice al banco de voz con el único fin de vender su producto. A partir de aquí, las voces virtuales que se vendan adoptarán la misma fórmula para llegar a más personas. Sin embargo, cabe resaltar que esto sucede en oriente y la estética que se utiliza en el apartado visual es propiamente de Japón. Es por este motivo que la gran mayoría de voces virtuales son de habla japonesa, al igual que el apartado artístico en los videos promocionales. Esta contextualización permite explicar el motivo por el cual hay un breve porcentaje del total de encuestados que no consideraría escuchar más música que incluya voces virtuales o también, no escuchar música que esté en japonés. Pese a que la cultura del “anime” se vea situado en el campo del entretenimiento, algunos de los oyentes pueden no tener preferencia por este estilo, sumando el hecho que se encuentre en japonés, ya que se asocia inmediatamente con la cultura japonesa. Las personas recuerdan y tienden a recordar una canción porque son memorables y las personas se sienten identificadas con la letra. En el caso de las canciones en un idioma que un individuo considere foráneo, lo único memorable es la melodía y demás elementos rítmico-musicales. La letra, en este caso, no tendría valor alguno y, por consiguiente, las canciones con un idioma foráneo no tendrían el mismo impacto que una canción en el idioma natal del público al que va dirigido.

Como se ha podido ver, son varios los elementos por los que los encuestados pueden no tener la misma preferencia por el uso de los cantantes virtuales. No obstante, esto no indica que el impacto haya sido nulo ya que cabe resaltar que, en primer lugar, los encuestados prefieren una voz “realista”. Esto puede contradecirse cuando los encuestados también prefieren canciones como la canción #2, siendo esta canción menos “realista” debido a los efectos aplicados en la voz. En este caso, se tiene que contextualizar qué es lo que permite que la música suene “real” para los encuestados. Lo “real” es sinónimo de lo que es lo natural dentro de su propio estilo. Un estilo de música que busca situar un entorno más

“digitalizado”, por ejemplo, va a sonar más congruente con el estilo que se adopte. Bajo esta premisa, una voz virtual no pretende ser un reemplazo para las personas que se dedican al canto, sino un recurso para potenciar un estilo musical en concreto. Esto explica el por qué la mayoría de encuestados considera que los cantantes virtuales funcionan como una herramienta más para los compositores. Esto quiere decir que el uso del *software* VOCALOID en la composición de canciones si tiene un impacto considerable en las personas encuestadas y también, por consiguiente, en los oyentes.



Conclusiones

La música que uno suele escuchar ha ido evolucionando a lo largo del tiempo. Y es por ello que, tanto la tecnología como la música, llegará un punto en la historia en que ya no importe cuestionarse si la música que uno escucha es “real” o “virtual”. Como bien menciona Collins en su artículo “Trading Faures: Virtual Musicians and Machine Ethics”, mientras más uno se expone por músicos virtuales, más raro y fantástico sería la vida musical (2011, p. 4). Lo que uno considera como “real” y lo que es “humano” es un concepto subjetivo que se asocia adaptándose a la realidad de cada individuo. Llegar a ello requiere comprender la realidad que engloba una cultura, un mundo con costumbres que difieren del que uno está acostumbrado, de lo que es “común”. El impacto que las personas pueden adoptar hacia un personaje que no es “real”, existe porque es asociada con algo, y ese “algo” está ligado a las tendencias que forma la realidad de su comunidad. Las primeras canciones compuestas con VOCALOID han permitido desarrollar canciones con un amplio rango vocal más de lo que un ser humano podría entonar. Esta característica es lo que ha permitido darle un toque propio al desarrollo de los temas musicales. Sin embargo, aún no se habla de la personalidad que adquiere el carácter de la voz virtual basándose en su apariencia gráfica. Esto se refleja en la temática que avalan las canciones y una creatividad que ponen como personaje principal a la propia cantante virtual como protagonista como se pudo ver en el *videoclip* de la canción #1.

Como se ha mencionado antes, para la mayoría de personas el “realismo” es importante en la originalidad con la que se utilice una voz virtual en una canción. Sin embargo, también es importante el *videoclip* para llegar a la audiencia. El estilo estético que se adopte para realizar las animaciones, transiciones y demás elementos audiovisuales tiene influencia en la idea que transmite el producto hacia el consumidor, tomando en cuenta que las canciones son en su gran mayoría en el idioma japonés. La música se ha convertido en

parte de la experiencia que las personas consumen, siendo esta “experiencia” una consecuencia del “producto” al complementarse con el *videoclip*. Ciertamente es una forma diferente de percibir la música ya que se cuestiona si los oyentes la “escuchan” o, en todo caso, la “observan”. A través de los videos promocionales que buscan publicitar la canción se establece la audiencia a la que está dirigida. Esto explica porqué hay mayor preferencia por el elemento realista, que corresponde al elemento auditivo en las canciones creadas con VOCALOID, pero los encuestados han preferido la canción #2 que es menos realista en cuanto al timbre de la voz virtual pero con un video promocional que envuelve una estética diferente que la canción #1, estableciendo el elemento visual de las canciones compuestas con VOCALOID. Este es el impacto que tiene los videos promocionales con relación a las canciones creadas con VOCALOID hacia los oyentes.

El elemento del “realismo” se asocia con la similitud que tiene un timbre de voz virtual comparado con un timbre de voz real. Sin embargo, lo “real” está asociado con lo que complementa el estilo de la canción. En sus inicios, el uso de las canciones creadas con VOCALOID ha sido relacionado con intentar parecerse a lo que puede hacer un cantante real, es decir, no ser capaz de encontrar su propio estilo. Sin embargo, el uso de VOCALOID con elementos que refuercen la estética de una canción ya sea por medio de *videoclips* y del estilo musical que adopte el tema musical, el uso de VOCALOID debe aportar para complementar el estilo musical para así ser parte del “producto” para así definir el realismo de la voz virtual. Es por eso que los encuestados prefieren canciones como la canción #2. La experimentación con el uso de este tipo de *softwares* no solamente exige que el compositor tiene que saber qué es lo que desea plasmar, sino que también aporta con nuevos recursos para la creación utilizando el *software* como una herramienta más de composición musical. VOCALOID es un programa que ofrece múltiples parámetros para que el compositor pueda plasmar sus creaciones con voz sintetizada a su manera sin requerir un cantante bajo las reglas que

engloban al género musical de la obra que se está componiendo. No necesariamente lo más ligado a lo “realista” es lo más estandarizado, sino la percepción del público frente al género en que se utilizan estas herramientas musicales ya que VOCALOID exprime las capacidades de una voz “irreal” para que el compositor tenga la libertad de utilizarlo en un contexto sonoro que esté alineado con su estética musical.



Referencias bibliográficas

- 1st Place Co., LMTD (2012) IA PROJECT [Página web]. <https://1stplace.co.jp/>
- Anselmo Del Pino, C. (2022). Decuantización, maximalismo y errores forzados: Análisis de los tres códigos dominantes de la música posdigital mediante la extracción de sus principales factores prácticos de codificación. *Inmaterial. Diseño, Arte y Sociedad*, 7(14), 117-143. <https://inmaterialdesign.com/INM/article/view/154>
- BACK-ON (2015). DOUKEI [Canción]. En *IA/03 –VISION-*. 1st Place Co. <https://open.spotify.com/intl-es/track/22ruexhkgf608ygFHp4dky?si=32ca93308e024046>
- BBC News (2012). MIDI, la tecnología que le abrió la puerta a la música digital. https://www.bbc.com/mundo/noticias/2012/12/121205_midi_tecnologia_musica_np
- Bell, S. (2016). The dB in the .db: Vocaloid Software as Posthuman Instrument. *Popular Music and Society*, 39(2), 222-240. <https://doi.org/10.1080/03007766.2015.1049041>
- Collins, N. (2011). Trading Faures: Virtual Musicians and Machine Ethics. *Leonardo Music Journal*, 21, 35-39. <https://www.jstor.org/stable/41416821>
- Cuevas, M. (2013). *Reportaje audiovisual VOCALOID: los cantantes del futuro* [Tesis de licenciatura, Universidad Nacional Autónoma de México]. Facultad de Ciencias Políticas Sociales. <https://goo.su/pB3GhQU>
- Glowacka, D. (2004). La música y su interpretación como vehículo de expresión y comunicación. *Comunicar. Revista Científica de Comunicación y Educación*, 23, 57-60. <https://www.revistacomunicar.com/verpdf.php?numero=23&articulo=23-2004-10>
- Goboumen (2012). Genjitsuteki Ronri Shugisha [Canción]. En *IA SUPERBEST – THE CREATORS-*. <https://open.spotify.com/intl-es/track/1bPUpRISzyfJFbvcNGepRX?si=f7b4f4b7dec3491e>

- Iserte, J. (2014). *Síntesis de voz cantada y canto coral: criterios musicales y estadísticos* [Trabajo de fin de grado, Escuela Superior Politécnica UPF]. Repositorio Digital de la UPF. <https://repositori.upf.edu/handle/10230/22885>
- Jorgensen, S. M. H., Vitting-Seerup, S., & Wallevik, K. (2017). Hatsune Miku: an uncertain image. *Digital Creativity*, 28(4), 318–331. <https://doi.org/10.1080/14626268.2017.1381625>
- Kenmochi, H. (25-30 de marzo de 2012). *Singing Synthesis as a New Musical Instrument* [Conferencia]. IEEE International Conference on Acoustics, Kyoto, Japón. <https://n9.cl/ey788>
- Klein, E. (2016). Feigning Humanity: Virtual Instruments, Simulation and Performativity. *Journal of the International Association for the Study of Popular Music*, 6(2), 22-43. <https://www.iaspmjournal.net>
- Mazzola, G. et al. (2018). New Means of Creation. En G. Mazzola et al., *Basic Music Technology: An introduction* (pp. 177-186). Springer. https://doi.org/10.1007/978-3-030-00982-3_22
- McLeod, K. (2016). Living in the immaterial world: Holograms and spirituality in recent popular music. *Popular Music and Society*, 39(5), 501-515. <https://goo.su/pN8Sy>
- Roseboro, B. (2019). *The vocaloid phenomenon: A glimpse into the future of songwriting, community-created content, art, and humanity* [Trabajo escolar, DePauw University]. Honor Scholar Thesis. <https://acortar.link/sp321h>
- Wilkinson, S. (2003). Humanoid or Vocaloid? *Electronic Musician*, 19(9), 28. <https://goo.su/jP0z8O>
- Yamaha Corporation. (2016). Vocaloid. <http://www.vocaloid.com/us/>
- Yin, Y. (2018). Vocaloid in China: Cosmopolitan music, cultural expresion, and multilayer identity. *Global Media and China*, 3(1), 51-66. <https://acortar.link/dYUida>