

VOX ET MACHINA. Apuntes para una genealogía de los usos y estéticas de la voz artificial en el pop

VOX ET MACHINA. Notes towards a genealogy of the uses and aesthetics of the artificial voice in pop



Jordi Roquer González

Universidad Autónoma de Barcelona, España

jordi.roquer@uab.cat

Resumen: Este artículo propone una genealogía parcial de los usos y estéticas de la voz artificial en la música popular. A partir de tecnologías pioneras como el Voder y el Vocoder, el texto recorre la progresiva incorporación de voces artificiales y procesadas en la música popular, desde las prácticas experimentales de la década de 1960 hasta su consolidación en la música electrónica, el pop y los géneros urbanos entre finales del siglo XX y principios del XXI. A través de casos que abarcan desde Kraftwerk y Laurie Anderson hasta Cher, Radiohead o Daft Punk, el artículo analiza cómo estas tecnologías vocales evolucionan progresivamente desde la novedad técnica hacia dispositivos expresivos, culturales y simbólicos. Lejos de concluir con un modelo teórico cerrado, la parte final desarrolla una síntesis interpretativa abierta, derivada del recorrido histórico y tecnológico planteado a lo largo del texto, y propone cinco posibles funciones de la voz artificial vinculadas a dimensiones estéticas, expresivas, temporales, técnicas y culturales. El artículo concluye con una coda sobre la inteligencia artificial y la clonación vocal, en la que se abordan los desafíos estéticos, legales y culturales que plantean los sistemas actuales de generación sintética de voz.

Palabras clave: voz artificial; música popular; vocoder; Auto-Tune; inteligencia artificial.

Abstract: This article proposes a partial genealogy of the uses and aesthetics of the artificial voice in popular music. Beginning with pioneering technologies such as the Voder and the Vocoder, the discussion traces the gradual incorporation of artificial and processed voices into popular music, from the experimental practices of the 1960s to their consolidation within electronic music, pop, and urban genres during the late twentieth and early twenty-first centuries. Through cases ranging from Kraftwerk and Laurie Anderson to Cher, Radiohead, and Daft Punk, the article examines how vocal technologies progressively evolve from technical novelty into expressive, cultural, and symbolic devices. Rather than concluding with a closed theoretical model, the final section develops an open interpretative synthesis derived from the historical and technological trajectory explored throughout the article, proposing five possible functions of the artificial voice linked to aesthetic, expressive, temporal, technical, and cultural dimensions. The article concludes with a coda on artificial intelligence and vocal cloning, addressing the aesthetic, legal and cultural challenges posed by contemporary systems of synthetic voice generation.

Keywords: artificial voice; popular music; vocoder; Auto-Tune; artificial intelligence.

Enviado en: 28 de enero de 2026

Aceptado en: 7 de junio de 2026

Publicado en: julio de 2026

Introducción

El presente artículo utiliza el término voz artificial para referirse a aquellas manifestaciones vocales cuya producción y/o transformación depende de tecnologías que modifican de manera sustancial su timbre, altura o inteligibilidad, desplazándolas del marco perceptivo asociado convencionalmente a la naturalidad vocal. Con todo, es necesario reconocer que esta distinción es relativa, dado que toda voz grabada, amplificada o reproducida técnicamente incorpora un grado inevitable de mediación tecnológica. El uso del término en este trabajo apunta, por tanto, a intervenciones aplicadas de forma deliberada para alterar el carácter orgánico de la voz con fines estéticos, expresivos o funcionales, generando configuraciones sonoras que tensionan los presupuestos tradicionales de autenticidad, corporalidad y performatividad vocal. Desde esta premisa, el artículo examina la evolución histórica, procedimental y estética de la voz artificial en la música popular. Tras una contextualización inicial, se propone una diacronía que abarca desde los primeros experimentos militares y científicos en síntesis de voz hasta su despliegue contemporáneo en sistemas basados en inteligencia artificial, pasando por los pioneros de la electrónica y su posterior integración en los lenguajes del pop y el rock. Finalmente, el texto propone una reflexión abierta sobre cinco posibles funciones de la voz artificial, entendidas no como categorías cerradas sino como síntesis interpretativas derivadas del recorrido histórico y tecnológico previamente planteado.

Antecedentes de la voz sintética

La aspiración de dotar de habla a objetos no humanos parece constituir un deseo recurrente en la imaginación técnica y cultural de múltiples civilizaciones. La historia de los autómatas musicales ofrece indicios tempranos de este

anhelo, como la referencia a una orquesta mecánica atribuida al emperador chino Han Gaozu en el siglo III a.C. (Needham, 1986, p. 158), la existencia de arpas eólicas en la Grecia clásica (Ord-Hume, 2004, p. 58) o el uso de mecanismos hidráulicos capaces de automatizar la emisión sonora (Cowell, 1980, p.160-162). Aunque estos ejemplos se sitúan todavía en el ámbito instrumental, revelan un marco simbólico fundacional donde lo mecánico y lo sonoro se articulan en torno a la idea de animación técnica de la materia.

El tránsito hacia una concepción científica de la voz artificial emerge en el contexto de la fisiología experimental del siglo XVIII. Christian Gottlieb Kratzenstein (1723-1795) consiguió reproducir vocales mediante tubos de resonancia acoplados a tubos de órgano (Schroeder, 1993, p. 231-237), demostrando que la voz podía concebirse como fenómeno susceptible de descomposición y replicación acústica. En paralelo, Wolfgang von Kempelen (1734-1804) —creador del célebre ajedrecista conocido como El Turco— avanzó posteriormente hacia una investigación genuinamente fonética con la construcción de su máquina parlante, documentada en *Mechanismus der menschlichen Sprache nebst Beschreibung einer sprechenden Maschine* (1791).¹ Esta obra se considera un punto de inflexión en la historia de la generación artificial del habla (Dudley y Tarnoczy, 1950. p. 151), situando la voz no ya como prodigio escénico sino como fenómeno replicable mediante control mecánico del flujo de aire, la articulación y la resonancia (Hankins y Silverman, 1995). A lo largo del siglo XIX surgen diversos dispositivos destinados a reproducir mecánicamente la voz humana, como la Euphonia de Joseph Faber, que confirma la posibilidad de generar habla mediante control artificial del flujo de aire y la resonancia (Lindsay, 1997). En esta trayectoria histórica resultan igualmente ineludibles las aportaciones de Hermann von Helmholtz (1819-1905), cuyos estudios sobre resonancia,

¹ La obra se publicó originalmente en francés y alemán de manera simultánea. Desde 2022 existe una traducción inglesa bajo el título *The Mechanism of Human Speech* (Wipf and Stock Publishers) que aporta información técnica adicional, así como algunas referencias bibliográficas omitidas en la versión original de 1791.

armónicos y fisiología auditiva establecen un marco conceptual crucial para comprender la estructura espectral de la voz y, por extensión, su posible reconstrucción artificial. A ello se suman las contribuciones de Melville Bell (1819-1905) y Alexander Graham Bell (1847-1922), que profundizan en la fonética y anatomía vocal, consolidando así la base científica necesaria para el estudio tecnoacústico de la producción del habla. Sobre estos fundamentos se asienta, ya en el siglo XX, el desarrollo de la síntesis electrónica del habla y posteriormente los sistemas digitales que desembocan en los modelos contemporáneos basados en inteligencia artificial.

Un viaje a los albores de la voz sintética: las tecnologías pioneras

El interés social por la robotización de la voz humana no adquiere verdadera relevancia estética hasta la década de 1930, con la aparición de dispositivos como el Sonovox y el Vocoder. Ambos constituyen hitos tempranos en la historia de la voz artificial, aunque responden a principios técnicos distintos: mientras el Vocoder analiza y reconstruye electrónicamente la voz, el Sonovox utiliza el tracto vocal humano como resonador de sonidos generados externamente, anticipando parcialmente el funcionamiento posterior del Talkbox. Una de las primeras apariciones cinematográficas documentadas la encontramos en *You'll Find Out* (dir. David Butler, 1940), donde Harry Babbitt interpreta "I'd Know You Anywhere" junto a la orquesta de Kay Kyser. En esta secuencia, la modulación vocal del sonido electrónico permite una ilusión performativa: la voz adopta el timbre de diversos instrumentos al aproximarse a ellos, escenificando la transferencia sonora entre máquina, cuerpo e instrumento musical. Más allá de la espectacularidad del efecto, este ejemplo ilustra cómo el Sonovox introduce en la cultura audiovisual una estética de mediación vocal explícita, en la que la artificialidad no se oculta, sino que se exhibe como signo sonoro distintivo.

Jacob Smith (2008) argumenta que parte del éxito del Sonovox radica precisamente en esa espectacularización, que posibilita “hacer hablar objetos” y lo convierte en una herramienta particularmente atractiva para la publicidad radiofónica y cinematográfica.² Su uso en jingles y anuncios consolida un imaginario en el que la voz artificial opera no solo como novedad tecnológica, sino como dispositivo retórico capaz de conferir agencia vocal a artefactos y productos, inaugurando un campo de experimentación que será central en la cultura mediática de mediados del siglo XX.

El Vocoder, desarrollado por Homer Dudley en los Bell Telephone Laboratories entre 1928 y 1939, supone un cambio decisivo al trasladar la voz al dominio del procesamiento electrónico. A diferencia del Sonovox —que utiliza el tracto vocal humano como resonador y filtro acústico— el Vocoder descompone la señal mediante filtros que analizan las envolventes formánticas, para luego reconstruirla sobre una fuente portadora artificial. Su primera presentación pública tuvo lugar en la Exposición Universal de Nueva York de 1939 con el Voder, interfaz operada manualmente que permitía articular sílabas y palabras mediante teclados y palancas. Más que imitación naturalista, el aparato ofrecía una voz mecánica audible, áspera y claramente artificial, inscrita en un imaginario tecnocientífico que permitía asociar el habla sintética con modernidad, eficiencia y control.

Tanto el Voder como el Vocoder fueron desarrollados por Homer Dudley en los Bell Telephone Laboratories en el marco de las investigaciones sobre transmisión de voz a larga distancia y optimización del ancho de banda. El objetivo no era inicialmente musical, sino tecnocientífico: descomponer la señal vocal a un conjunto mínimo de parámetros articulatorios que permitieran reconstruir el habla tras la transmisión, separando el contenido lingüístico de su materialidad acústica.

² Smith, Jacob. “Tearing speech to pieces: voice technologies of the 1940s.” *Music, Sound, and the Moving Image*, vol. 2, no. 2, Dec. 2008, p. 183.

Esta lógica funcional se radicalizó durante la Segunda Guerra Mundial con el proyecto SIGSALY (1943), primer sistema de comunicación digital y encriptada de la historia, en el que el Vocoder se integró como módulo de síntesis y análisis para asegurar conversaciones seguras entre altos mandos aliados, como Winston Churchill y Franklin D. Roosevelt (Wheatley, 1953). La voz se convierte así en información cuantificable, susceptible de descomposición matemática y recombinación electrónica: un cambio epistemológico decisivo que inaugura la posibilidad de pensar la voz como dato y no únicamente como gesto corporal.

Tras el conflicto bélico, el Vocoder se adapta a aplicaciones civiles y pronto encuentra un lugar destacado en la cultura audiovisual, especialmente en la creación de voces que evocan tecnologías futuristas o inteligencias artificiales. Aunque sus primeras apariciones cinematográficas son todavía puntuales y experimentales, su imaginario se consolida a partir de los años sesenta, cuando la voz sintética pasa de ser un experimento de laboratorio a un recurso retórico y estético. Un ejemplo paradigmático se encuentra en *2001: A Space Odyssey* (dir. Stanley Kubrick, 1968), donde la voz de HAL 9000 —interpretada por Douglas Rain sin procesamiento electrónico extremo— encarna la paradoja de la voz artificial que suena demasiado humana, inquietante precisamente por su exceso de naturalidad. La célebre escena en la que HAL es desactivado mientras canta la canción “Daisy Bell”, remite directamente a la primera demostración documentada de canto computacional realizada en 1961 con un IBM 7094 en los laboratorios Bell. En aquella prueba histórica, John L. Kelly Jr. y Carol Lochbaum sintetizaron la voz formántica, mientras que Max Mathews programó el acompañamiento musical. Entre los asistentes se encontraba Arthur C. Clarke, quien años más tarde desarrollaría junto a Stanley Kubrick el universo narrativo de *2001: A Space Odyssey*, donde “Daisy Bell” reaparece resignificada como emblema de regresión cognitiva mecánica y de la fragilidad de la supuesta autonomía de la máquina.

Otros usos del Vocoder consolidaron este imaginario en la cultura popular: la reinterpretación del “Himno a la Alegría” realizada por Wendy Carlos para *A Clockwork Orange* (dir. Kubrick, 1971) proyectó la voz sintetizada hacia una estética protoelectrónica de distanciamiento emocional, mientras que “The Robots” (Kraftwerk, EMI, 1978) formalizó la iconografía *cyborg* como performance sonora y visual. La figura del intérprete híbrido —mitad cuerpo, mitad circuito— se alejaba de la especulación literaria para convertirse en estética reconocible y reproducible. La robotización vocal se convierte así en un marcador sonoro de modernidad tecnológica, un significativo cultural antes que un mero efecto de procesado. Retomaremos más adelante el Vocoder para recalcar en su desarrollo musical y en sus implicaciones estéticas en la segunda mitad del siglo XX. Antes, conviene atender a otros artefactos clave que emergen en las décadas de 1960 y 1970, y que ampliarán este panorama desde perspectivas centradas en la modulación corporal y la performatividad escénica de la voz artificial. Como argumenta Brøvig-Hanssen (2010), en los dispositivos de alteración de la voz la presencia vocal nunca desaparece, sino que se vuelve perceptible como parte de un efecto que tiende a generar fricción estética. Esta observación resulta particularmente relevante a la luz de la distinción entre mediación transparente y mediación opaca formulada por la propia autora: en una mediación transparente, la tecnología se integra de forma que no interfiera perceptivamente en la escucha, permitiendo que el resultado sonoro se perciba como naturalizado o no mediado. Aunque el efecto está ahí, no es percibido por el público. Por el contrario, la mediación opaca hace audible el proceso técnico, la intervención de la tecnología no queda oculta, sino que se manifiesta como parte del timbre y de la textura resultante. Aunque este marco no fue propuesto específicamente para la voz, sino para comprender cómo las tecnologías influyen en la percepción sonora y la producción musical en general, resulta pertinente para interpretar la fricción estética y la recepción social de la voz artificial. Más

adelante, este mismo prisma permitirá abordar casos en los que la tecnología vocal —del Vocoder al Auto-Tune o la síntesis con IA— oscila entre estrategias de opacidad perceptible y procesos que, con el tiempo, pueden devenir transparentes por naturalización cultural.

Experimentación y transformación

La década de 1960 supone un punto de inflexión decisivo en la relación entre tecnología y producción musical. Aunque muchas de las herramientas implicadas —grabación multipista, manipulación de cinta, reverberación artificial, efectos de modulación y síntesis electrónica— ya habían sido exploradas previamente en el ámbito de la música concreta y electroacústica, este es el momento en que varias de ellas se incorporan de manera estable a la música popular. El estudio de grabación deja así de concebirse como un espacio neutral destinado a capturar una realidad sonora preexistente para convertirse en un instrumento creativo activo. La evolución de las tecnologías de captación sonora —especialmente el desarrollo de micrófonos de alta sensibilidad y las técnicas de grabación cercana— favoreció nuevas formas de vocalidad íntima y detallista que transformarían progresivamente los modos de cantar en la música popular. Sobre este nuevo marco técnico se consolidan, durante los años sesenta y setenta, prácticas de manipulación y procesamiento vocal cada vez más creativas, desde la modificación espacial y tímbrica hasta la alteración deliberada de parámetros considerados tradicionalmente “naturales”. En este contexto nacen interesantes experimentos a partir del uso de *phasers*, *flangers* y otros efectos basados en la modulación de la señal, como por ejemplo el denominado ADT (*Automatic Double Tracking*) célebremente usado por John Lennon desde la época de los Beatles. En esa técnica desarrollada por Ken Townsend en Abbey Road, la señal vocal se duplicaba de forma automática mediante un ligero desfase temporal controlado, evitando la necesidad de grabar una segunda toma

y generando una sensación de grosor y presencia que pronto se convirtió en un recurso muy usado en el pop *mainstream*. Se trata de un claro ejemplo de mediación transparente: el oyente apenas percibe el efecto que queda oculto entre la sutilidad y la mediación tecnológica general propia de la época.

En el extremo opuesto situamos el Talkbox, cuyo uso se intensifica a finales de los sesenta y especialmente en los años setenta. A diferencia del Vocoder —donde la señal vocal se analiza y sintetiza electrónicamente— el Talkbox remite a un principio físico más directo: el sonido del instrumento (guitarra, teclado u otros) se envía a través de un tubo a la boca del intérprete, que modula la señal mediante movimientos de labios, lengua y cavidad bucal. El resultado recuerda al Sonovox de 1940, pero aquí la dimensión corporal y gestual adquiere un papel determinante. Mientras que en el Vocoder la mediación tecnológica permanece desligada del cuerpo del intérprete, en el Talkbox la tecnología se incorpora físicamente a la performance. El tubo visible, la acción bucal y la proximidad del artefacto al instrumento convierten el procesamiento vocal en un gesto escénico explícito. Esto altera radicalmente la recepción del artificio: no hay engaño, sino una ampliación performativa del cuerpo musical. La voz híbrida humano-máquina deja de percibirse como suplantación o artificio encubierto y pasa a entenderse como una prolongación tecnológica del intérprete. El caso más paradigmático es “Do You Feel Like We Do?” de Peter Frampton (A&M Records, 1976), cuando, en directo, el uso del Talkbox se transforma en un momento ritual compartido: el público reconoce el artefacto, valida el artificio y lo incorpora como experiencia interactiva.³ La tecnología ya no se sitúa necesariamente en tensión con la idea de autenticidad performativa, sino que puede incluso reforzarla al integrarse en el cuerpo y el gesto del músico.

³ En las interpretaciones en directo de la pieza, Frampton suele utilizar el Talkbox para formular al público la pregunta “Do you feel like we do?”, convertida ya en estribillo colectivo. La audiencia responde coreando la frase, integrando así el artificio tecnológico dentro de una dinámica participativa y ritualizada muy propia del denominado “rock de estadio”.

Pete Drake es frecuentemente citado como pionero en la utilización del Talkbox en un registro comercial, especialmente en el álbum *Forever* (Stop Records, 1964), donde el *pedal steel* asume cualidades vocales difícilmente asociables a un instrumento tradicional. La técnica fue adoptada posteriormente en el rock y el funk, con ejemplos bien documentados como Joe Walsh en *Rocky Mountain Way* (ABC Records, 1973) y el ya mencionado Peter Frampton, no solo en "Do You Feel Like We Do?" sino también en otros éxitos comerciales como "Show Me the Way" (A&M, 1976). Años más tarde, Richie Sambora contribuirá a consolidar su presencia estética en el rock popular a través de temas como "Livin' on a Prayer" (Mercury, 1986) o "It's My Life" (Island, 2000).

Un ejemplo especialmente revelador es el uso del Talkbox por Stevie Wonder. A diferencia de Drake o Frampton, Wonder emplea el dispositivo con una sensibilidad melódica y una dimensión afectiva inusuales, introduciendo una matización expresiva que excede lo meramente efectista. Su interpretación en directo de "Close to You" de los Carpenters en The David Frost Show (1972) muestra una articulación vocalizada del Talkbox que progresa hacia lo íntimo y lo melancólico, abriendo un registro que anticipa futuros usos más emocionales de la voz tecnológicamente mediada. Wonder utiliza el Talkbox de manera más integrada en una interpretación tradicional. El control preciso del tracto vocal le permite modular la señal con suficiente flexibilidad como para aproximarla a ciertos matices de la voz cantada convencional, tanto a nivel dinámico como tímbrico. El resultado no elimina completamente la percepción del dispositivo, pero sí desplaza la atención desde el efecto tecnológico hacia la expresividad de la interpretación. La voz procesada deja así de percibirse únicamente como artificio externo y pasa a incorporarse al propio gesto performativo del músico.

De los ochenta a los noventa: la consolidación cultural de la voz artificial

En los años ochenta la voz artificial deja de ser un recurso excéntrico para instalarse en determinadas corrientes *mainstream* de la cultura popular. En el hip-hop y la música electrónica, canciones como “Planet Rock” de Afrika Bambaataa (Tommy Boy, 1982) o “Jam on It” de Newcleus (Sunnyview Records, 1984) muestran cómo la voz electrónicamente mediada puede encarnar un futurismo urbano vinculado al nacimiento del *breakdance*, la cultura *block party* y el rol del DJ como nuevo mediador tecnológico de la voz y del cuerpo. En este contexto, el timbre robótico funciona como emblema de modernidad y, al mismo tiempo, como signo de pertenencia generacional. La voz sintetizada acompaña al *beat* y al *scratch*, integrándose en la iconografía visual y sonora de la cultura hip-hop emergente.

También el pop experimental encuentra en la voz artificial una herramienta crítica. Laurie Anderson, en “O Superman” (Warner, 1981), utiliza un procesador vocal en combinación con un patrón rítmico minimalista que refuerza el carácter hipnótico y despersonalizado del mensaje. La frialdad de la voz tratada subraya la ironía del texto –“O Superman, O Judge, O Mom and Dad”– y sugiere que el artificio tecnológico también puede funcionar como mediador de nuestras relaciones más íntimas. En “Oh Superman”, el tratamiento vocal —basado en un bucle rítmico mínimo y una modulación electrónica estable— elimina gran parte de la variabilidad prosódica asociada a la voz humana: la entonación se aplanan, el fraseo se regulariza y la dinámica queda contenida dentro de un rango estrecho. Esta monotonía no es un efecto colateral, sino una estrategia compositiva basada en el propio efecto. La relación entre voz, ritmo y *loop* resulta aquí central: la repetición insistente de la sílaba “ha” no funciona como elemento melódico en sentido tradicional, sino como estructura temporal que sostiene el discurso. La ausencia de inflexión emocional no neutraliza el mensaje, sino que lo intensifica, generando una distancia que

refuerza la dimensión crítica del texto. No obstante, esta lectura admite matices: la aparente frialdad puede interpretarse también como una forma de afectividad mediada, donde la contención expresiva no implica ausencia de emoción, sino su desplazamiento hacia el plano estructural. De nuevo, como en el caso de Wonder, es precisamente en esta ambigüedad entre automatismo y agencia discursiva, donde el procesamiento vocal se convierte en elemento central del significado.⁴

También en el rock de los ochenta se exploran dimensiones inusitadas de la voz procesada. Neil Young recurre al Vocoder en varias canciones de *Trans* (Geffen, 1982), en un contexto profundamente marcado por los intentos de comunicación con su hijo Ben, afectado por parálisis cerebral. El propio Young explicaría años más tarde que la dificultad para comprender las voces procesadas del álbum reflejaba parcialmente la experiencia comunicativa que vivía entonces con su hijo.⁵ Este gesto personal se puede leer en clave de resignificación de la tecnología: lo que en otros contextos evoca frialdad o alienación, en *Trans* aparece como puente afectivo, como mediación empática ante la imposibilidad del habla verbal.

Las últimas décadas del siglo XX muestran, por tanto, la extraordinaria versatilidad semántica de la voz artificial. Las voces procesadas pueden desplazarse entre lo comercial (como en "Mr. Blue Sky" de Electric Light Orchestra), lo futurista (Afrika Bambaataa), lo alegórico y crítico (Anderson) o lo íntimo y afectivo (Young), para integrarse posteriormente de forma normalizada en distintas músicas híbridas entre la electrónica y el pop, como ocurrirá más adelante en casos como el de Daft Punk. Esta diversidad señala un cambio fundamental: la voz artificial deja de ocupar un margen experimental para convertirse en un recurso consolidado, capaz de habitar simultáneamente el espectáculo masivo, la crítica social y la emoción personal. A finales de los noventa este proceso desembocará en la expansión del Auto-Tune y otros procesadores digitales, donde

⁴ Por ejemplo, Susan McClary (2001, p. 35-45) ofrece una lectura cultural de la obra en el marco de la performance multimedia y las poéticas del minimalismo.

⁵ Neil Young Archives. "Album of the Week: Trans". 12 jun. 2019, citado en: A Pop Life. Acceso: 16 mayo 2026.

la mediación ya no se exhibe visualmente como en el Talkbox, sino que opera de forma audible pero invisible, generando reacciones estéticas ambivalentes entre fascinación, sospecha y rechazo.

Narrativas de los noventa: entre el afecto y la alienación

Si el caso de Stevie Wonder abría la posibilidad de una expresividad melancólica en la voz artificial, en los años noventa esa dimensión se expande y adquiere un papel central en la construcción de narrativas musicales. Para ilustrarlo visitaremos el disco *Three Sides to Every Story* (A&M, 1992), del grupo Extreme. El cierre de este álbum corresponde a una suite formada por tres secciones interconectadas donde convergen motivos musicales, materiales textuales y gestos retóricos ya presentados a lo largo del disco. En la parte final, Gary Cherone entona una plegaria dirigida a Jesús en clave de confesión existencial sobre el sinsentido vital. La estrofa conduce a un estribillo dialogado donde la voz solista interactúa con un Vocoder que responde repetidamente “who cares..?”. La atribución instrumental en los créditos del álbum indica que, tras esa voz procesada, se encuentra Nuno Bettencourt; en escena, técnicamente, estaríamos ante un diálogo entre cantante y guitarrista. Sin embargo, y aquí emerge el interés musicológico, desde una escucha extradiegética, el efecto abre un horizonte interpretativo mucho más amplio. La deshumanización tímbrica del Vocoder desencadena lecturas múltiples: ¿habla el propio Jesús? ¿es una entidad divina o una voz interior? ¿un desdoblamiento psicológico? ¿una conciencia culpable que interpela al sujeto? En ausencia de timbre humano reconocible, el oyente imagina el origen de la voz. La alternancia entre pregunta/respuesta (Vocoder interpela – Cherone responde; luego Cherone pregunta – la voz artificial replica) produce un juego discursivo casi teatral, permitiendo visualizar dos agentes en diálogo.

Siguiendo con los noventa, resulta ineludible mencionar *OK Computer* (Parlophone, 1997) de Radiohead, obra clave para comprender los derroteros del rock alternativo a finales de siglo. En el corte “Fitter Happier”, aparece probablemente una de las primeras aplicaciones explícitas de *text-to-speech* en un disco *mainstream* (usando a Fred, la voz sintética de los Macintosh de la época). Thom Yorke ha explicado que resultó “liberador” dar esas palabras a una computadora para que las ejecutara con esa frialdad neutral.⁶ Desde el punto de vista analítico, esta restricción prosódica produce un efecto significativo: la ausencia de variación expresiva desplaza el peso semántico desde la interpretación vocal hacia la estructura textual. La voz no “expresa” el contenido, sino que lo expone, generando una distancia que intensifica su carácter crítico. No obstante, esta aparente neutralidad también puede interpretarse de forma ambivalente. Aunque la voz carece de rasgos emocionales reconocibles, su propia artificialidad introduce una capa de significado que el oyente no puede obviar: la sensación de extrañamiento no proviene solo del texto, sino de la disociación entre lenguaje y cuerpo. Según el propio Yorke, aquel texto funciona como lista de consignas para la década, una radiografía ácida de la condición humana contemporánea.⁷ Por su parte, Tim Footman, argumenta que el sujeto de la canción representa al “ser humano occidental moderno, cómodo materialmente, pero vacío moralmente”.⁸ A nivel musical, el uso del *score* de la película *Three Days of the Condor* (Dir. Sydney Pollack, 1975) intensifica la sensación de extrañamiento y paradoja emocional.

Si observamos el hilo histórico, la experimentación sonora de los sesenta y primeros setenta abrió un nuevo terreno donde la voz dejó de ser un índice exclusivo de autenticidad para convertirse también en superficie manipulable. Lo que comenzó como curiosidad técnica evolucionó hacia un lenguaje

6 Moran, Caitlin (1997). “I was feeling incredible hysteria and panic...”. Select: 92.

7 Sutherland, Mark (1997), “Return of the Mac!”, Melody Maker

8 Footman, Tim (2007). Welcome to the Machine: OK Computer and the Death of the Classic Album. New Malden: Chrome Dreams.

expresivo estable. Desde esta perspectiva, los ochenta y noventa representan un punto de inflexión: tecnologías como el Vocoder, el Talkbox y la síntesis de texto a voz entran en diálogo directo con la emocionalidad pop, amplificando tanto el registro afectivo como el alienante. La voz artificial abandona la marginalidad y pasa a ocupar posiciones centrales en el relato musical, preludiando los debates sobre identidad, autenticidad y artificialidad que marcarán la transición al nuevo milenio.

Hacia el siglo XXI: Auto-Tune y más allá

Tras las exploraciones expresivas de las décadas anteriores, la transición hacia el siglo XXI está marcada por una herramienta que transforma de manera decisiva la relación entre cuerpo, voz y tecnología. Nos referimos al ahora ineludible Auto-Tune, aparecido a finales de los años noventa y convertido en poco tiempo en uno de los dispositivos más influyentes de la música popular contemporánea, así como en el último gran catalizador del debate recurrente entre autenticidad y mediación tecnológica. El tema que pone el Auto-Tune a los oídos del gran público es, sin duda, "Believe" de Cher (Warner, 1998), convertido en un hito por el carácter audible y distintivo de su tratamiento vocal. Aunque Cher no fue la primera en emplearlo, sí logró situarlo en el centro del imaginario mediático, haciendo visible una técnica que hasta entonces había operado mayoritariamente como recurso correctivo discreto. Artistas como Roy Veda o Eiffel 65 ya habían experimentado con el efecto, aunque sin desencadenar la misma resonancia cultural. El caso de "Believe" marca un punto de inflexión en la historia reciente de la voz artificial, no solo por su impacto cultural, sino por la especificidad de su tratamiento sonoro. Aquí el Auto-Tune se configura con una velocidad de respuesta extremadamente alta, lo que elimina las transiciones continuas entre alturas (*portamenti*) y produce saltos abruptos entre notas discretas. Este efecto no corrige la voz, sino que la reconfigura, imponiendo una lógica cuantizada del *pitch* que transforma la continuidad

expresiva en segmentación. El timbre sigue remitiendo a una voz reconocible, pero el comportamiento melódico — excesivamente estable y carente de microvariaciones— rompe con las expectativas asociadas al canto humano. El resultado no es simplemente una voz “afinada”, sino una voz cuya expresividad ha sido parcialmente externalizada al algoritmo.

Resulta interesante observar cómo fue recibida esta sonoridad en su contexto original, en un ejercicio hemerográfico que ilustra la tensión generada por el nuevo efecto. En 1998, Michael Gallucci escribía sobre “Believe” en Ultimate Classic Rock:

Cher simplemente se mueve a través de una fábrica de ritmos con un rango vocal monótono, mezclando bailables y semi-baladas en una sesión interminable y sin personalidad. Cualquier cantante podría haber sido contratada para la voz principal para estas canciones (que a veces se deslizan hacia el plástico brillante y sobre producido de finales de los 80). Cher solo pone su nombre e influencia de estrella, sin alma alguna. (Gallucci, cit. en Lacasse, 2000). Trad. del autor.

A lo que Serge Lacasse, en su tesis sobre la voz en la música popular, añade:

Esta asociación, apenas disimulada, entre la presencia del efecto artificial y sobre producido y la inmutable “belleza plástica” de Cher eran, por supuesto, inevitables... Pero más allá de eso, fue el impacto del efecto en el público lo que llamó la atención. En el caso de Cher, tanto la novedad del efecto como su utilización por una artista famosa normalmente no asociada con este tipo de artificios tecnológicos, parecen ser la raíz de ese amplio interés (2000, p. 2). Trad. del autor.

Y Dave Tompkins, autor del volumen *The Vocoder from World War II to Hip-Hop*, se suma a la analogía con la cirugía plástica señalando que “el Auto-Tune popularizado por la ‘robotox’ de Cher [...] fue confundido por el Vocoder, dándole [a Cher] una buena coartada desde un falso revival” (2010, p. 29). La observación es significativa: el público asoció aquel timbre procesado con el Vocoder clásico, cuando lo que en realidad estaba escuchando era el despliegue público de una tecnología nueva que, en breve, transformaría los modos de cantar y producir voz en la música popular. Sin duda, pasadas más de dos décadas, su progresiva normalización en el pop y las músicas urbanas sugiere un desplazamiento hacia formas de mediación cada vez más transparentes (Brøvig-Hanssen, 2010; Rey, 2020, 2024), donde la intervención tecnológica deja de percibirse como anomalía para integrarse en el estándar vocal contemporáneo.

Poco después de “Believe”, y recuperando bajo un nuevo prisma el binomio hombre-máquina que Kraftwerk había iconizado en los setenta, emerge la estética Vocoder-electrónica de Daft Punk. Con “Harder, Better, Faster, Stronger” (Virgin, 2001), el dúo francés propone una arquitectura vocal basada en el recorte y la reordenación digital, convirtiendo la voz en una serie de módulos permutables cuya inteligibilidad depende tanto del timbre como del montaje. Ese mismo gesto se expande en piezas como “One More Time” y “Digital Love” (ambas en *Discovery*, Virgin, 2001), y reaparece más de una década después en *Random Access Memories* (Columbia, 2013), donde canciones como “Instant Crush” desarrollan un tratamiento vocal fuertemente asociado al ADN sonoro del grupo. En su etapa final, colaboraciones con artistas como The Weeknd —“Starboy” y “I Feel It Coming” (Republic, 2016)— mantienen el Vocoder como marca identitaria indiscutible, algo que resulta especialmente interesante en un contexto de colaboración: incluso cuando no ocupa el eje frontal del timbre principal, el tratamiento aparece integrado como textura reconocible, permitiendo identificar al

dúo en la escucha. En términos de Askerøi (2013; 2016), funciona como marcador sonoro, al señalar una genealogía estilística vinculada al pop electrónico y al imaginario retrofuturista que Daft Punk hereda de Kraftwerk y reelabora para el siglo XXI. Al mismo tiempo, puede entenderse como *sound signature* en el sentido propuesto por Zagorski-Thomas (2014), es decir, una firma tímbrico-productiva que permite reconocer al artista aun sin mediación visual. En determinadas secciones y armonías, la voz procesada llega a operar como presencia vocal paralela, una forma de *vocal persona*⁹ que refuerza la percepción de diálogo o dúo incluso sin que el dúo aparezca en pantalla. Pero no se trata solo de una cuestión estilística, sino también de algo relevante en el plano de la recepción: el Vocoder (y el resto de artefactos tratados en este artículo) pueden leerse también como significantes sonoros en el sentido planteado por Tagg (2013), dado que su timbre es capaz de activar asociaciones culturales o afectivas como futurismo, deshumanización lúdica, tecno hedonismo o incluso retrofuturismo —ese futuro imaginado desde el pasado que retorna, estilizado, en pleno siglo XXI.

La superposición de estos marcos muestra algo esencial para nuestro objeto de estudio: un mismo fenómeno vocal puede simultáneamente significar, marcar identidad y firmar una estética. En otros términos, la voz artificial puede entenderse no solo como efecto u ornamento, sino vector cultural capaz de producir sentido. Así pues, en boca de Daft Punk —nunca mejor dicho— el Vocoder, el Talkbox y otros procesadores no intentan ocultar su artificio; lo teatralizan. El cuerpo retrocede, el circuito avanza, y la agencia vocal se redistribuye entre máquina, producción y performance. Esa revolución se consolida con artistas como T-Pain, quien convierte el Auto-Tune en su marca sonora durante la década de 2000, hasta el punto de ser identificado popularmente con el propio efecto. Kanye West,

9 El concepto de *vocal persona* procede también del marco semiótico de Philip Tagg, donde designa la identidad vocal construida en una grabación: no la voz física del cantante, sino su figura performativa mediada por producción, timbre y estilo (Tagg, 2013, p. 306-312).

con *808s & Heartbreak* (Def Jam Recordings, 2008), lo lleva a un territorio distinto: no como embellecedor vocal, sino como filtro de vulnerabilidad, dolor y melancolía. El Auto-Tune, que a finales de los noventa había sido recibido mayoritariamente como artificio llamativo, pasa en apenas una década a operar como medio legitimado de expresión emocional. En paralelo, el reguetón, el trap y otras músicas urbanas lo integran como rasgo estilístico permanente, contribuyendo a que generaciones enteras crezcan percibiendo la voz procesada a través de una mediación cada vez más transparente.

Lo cierto es que, desde finales del siglo XX, la voz artificial ha dejado de ser recurso episódico para convertirse en lenguaje estructural. El recorrido histórico que hemos trazado permite plantear ahora la pregunta central de este artículo: ¿cuáles son las funciones expresivas, simbólicas y estéticas que la voz artificial ha desempeñado en la música popular? Más allá de tecnologías y ejemplos puntuales, proponemos entender su uso bajo un conjunto de intenciones operativas: cinco ejes que permiten identificar ciertas regularidades discursivas en su circulación sonora y cultural.

Cinco posibles funciones de la voz artificial: una síntesis interpretativa

Tras el recorrido histórico y tecnológico desarrollado hasta aquí, las funciones que se proponen a continuación deben entenderse no como una tipología cerrada o exhaustiva, sino como una síntesis interpretativa orientada a identificar ciertas regularidades en los usos culturales de la voz artificial dentro de la música popular. Estas funciones se articulan a partir de la interacción entre tres dimensiones: las transformaciones sonoras (cómo se modifica la voz), los efectos perceptivos (cómo se escucha) y las significaciones culturales que adquiere en contextos específicos. No describen tecnologías concretas, sino modos de uso que pueden coexistir, superponerse o

desplazarse dentro de un mismo caso. Se trata, por tanto, de un esquema interpretativo y flexible que permite pensar la voz artificial no solo como efecto técnico, sino también como práctica cultural cargada de agencia y sentido. A partir de ello, proponemos cinco grandes ejes funcionales que desarrollamos a continuación:

a) Transgresión estética y creativa

Esta función se identifica cuando la intervención tecnológica busca una ruptura con las convenciones de naturalidad o con marcos estéticos previamente estabilizados. Bajo ese prisma, puede entenderse como una forma de fricción estética, en la medida en que introduce tensiones entre expectativas de escucha y material sonoro, generando efectos de extrañamiento (Shklovsky, 1917)¹⁰ o desestabilización de códigos establecidos (Middleton, 1990). Todas y cada una de las herramientas comentadas (ya sea Auto-Tune, Vocoder, Talkbox o la síntesis de voz) tienen en común la transgresión sonora sobre un elemento tan íntimo y orgánico para el ser humano como es su propia voz. Y, más allá de la acción estética en lo sonoro, estas herramientas abren un universo de posibilidades para los creadores. El cine de la segunda mitad del siglo XX opera como un mediador indiscutible en la construcción social de una idea de voz artificial; determinadas músicas experimentales de los setenta (con discursos que penetran en el *mainstream*, como en el caso de Kraftwerk) trasladan esas tecnologías de lo audiovisual al terreno de la recepción musical. A finales del siglo XX y principios del XXI se observa cómo productores y artistas adoptan estas tecnologías no solo como efecto puntual, sino como parte integral de su estética. Tal expansión no se limita al timbre: afecta a la forma de componer, a la dramaturgia de los conciertos y a la percepción cultural de lo que entendemos

¹⁰ El concepto de extrañamiento (*ostranenie*) acuñado por Viktor Shklovsky (1917) entiende el arte como algo capaz de transgredir la percepción cotidiana. En el contexto de la música popular, esta capacidad de transgresión no solo opera a través de elementos tradicionales de la expresión musical sino también a través de la mediación tecnológica (Auto-Tune, síntesis, distorsión, etc.).

como “voz cantada”. Y esos cambios afectan también al régimen de escucha: donde antes buscábamos la correspondencia voz-cuerpo, ahora escuchamos voces que pueden ser múltiples, duplicadas, distribuidas o híbridas. La agencia vocal deja de residir únicamente en el intérprete y se reparte entre software, producción y ejecución.

No obstante, cabe señalar que el carácter transgresor de estas prácticas no es inherente ni estable, sino que depende de contextos históricos y culturales específicos, pudiendo ser rápidamente normalizado en determinados entornos de producción y escucha. De nuevo, tal razonamiento nos acerca al binomio entre mediación opaca y transparente, comentado ya en puntos anteriores. En este sentido, la transgresión debe entenderse como una condición situada, en ningún caso como una propiedad esencial de la tecnología.

b) Potencial expresivo

De la misma manera que a lo largo del siglo XX el micrófono generó nuevas maneras de cantar y nuevas sensibilidades en su recepción, la voz artificial también ha generado nuevos modelos de expresión emocional. Estas tecnologías han contribuido a que aflore un nuevo “grano de la voz” (término acuñado por Roland Barthes) que Simon Frith adapta para analizar las cualidades distintivas, a menudo subjetivas y emocionales, que dan carácter a una voz. Frith lo relaciona, por ejemplo, con el gusto por la voz quebrada en la música popular y su capacidad de generar vinculación emocional. El procesamiento digital permite imaginar un nuevo grano tecnológico, no orgánico sino mecánico-afectivo: una emoción filtrada que no nace de las cuerdas vocales, sino de la mediación técnica. Esta reformulación no implica la desaparición de la dimensión corporal de la voz (central para Barthes), sino su reconfiguración a través de la mediación técnica. Los artistas no siempre usan el Vocoder y el Auto-Tune para ocultar defectos, sino para modelar un discurso estético y expresivo. Artistas de reguetón

y trap han llevado el recurso a la exaltación hedonista, donde la voz afinada digitalmente puede transmitir deseo, poder y exceso. La expresividad vocal, por tanto, ya no depende exclusivamente del canto: también el artificio tecnológico se convierte en signo afectivo capaz de transmitir determinados mensajes. Esa deshumanización afectiva de la voz puede utilizarse de muchas formas, pero su mayor potencial reside en la creación de nuevos narradores, que implican la generación de nuevas sensibilidades. Lo hemos visto en los casos de Stevie Wonder, Extreme o en la dimensión crítica del mensaje de Radiohead a través de la voz robótica de Fred. Por lo tanto, podemos afirmar que la voz robotizada (especialmente el Vocoder) se ha convertido en un elemento icónico de la cultura pop contemporánea, como signo audible de no-humanidad emocional o afectividad mediada.

c) Modernidad y retromodernidad

El sonido del Vocoder puede asociarse simultáneamente con imaginarios tecnológicos del pasado y del futuro. Escuchado hoy, remite con frecuencia a determinadas estéticas electrónicas de las décadas de 1970 y 1980, pero continúa utilizándose también para construir atmósferas futuristas tanto en el ámbito audiovisual como en la música. Se trata de uno de los dispositivos sonoros más paradigmáticos de la cultura tecnológica moderna, precisamente por su capacidad para activar temporalidades múltiples y parcialmente contradictorias. No obstante, esta ambivalencia no es exclusiva del Vocoder, sino que puede observarse también en otros timbres y recursos vinculados históricamente a la imaginación electrónica. La paradoja se acentúa con fenómenos recientes como el *vaporwave* o el *synthwave*, géneros que recuperan imaginarios sonoros asociados a la cultura digital de los años ochenta mediante herramientas contemporáneas de producción. En estos contextos, el Vocoder deja parcialmente de percibirse como signo de futurismo para convertirse en marcador nostálgico de futuros imaginados desde el pasado. Esta condición

retrofuturista puede ponerse en relación con lo que Andreas Huyssen (2003) describió como la creciente centralidad cultural de la nostalgia en las sociedades contemporáneas, así como con las reflexiones de Simon Reynolds (2011) sobre la cultura pop obsesionada con su propio archivo y reciclaje estilístico. En una línea similar, Mark Fisher (2014) señaló cómo buena parte de la cultura contemporánea parece experimentar el futuro no como promesa abierta, sino como repertorio de posibilidades perdidas o no realizadas. Desde estas perspectivas, determinados timbres asociados históricamente a la tecnología —como el Vocoder— pueden adquirir una dimensión temporal ambigua: ya no representan simplemente “el futuro”, sino la memoria cultural de futuros imaginados en décadas anteriores. Podría plantearse tentativamente que ciertos usos contemporáneos de la voz artificial generan formas de percepción temporal híbrida, donde pasado y futuro aparecen superpuestos dentro de una misma experiencia sonora.

d) Asistencia o sustitución

Las distintas tecnologías discutidas permiten, en mayor o menor grado, transformar la materialidad sonora de la voz cantada, haciendo posible no solo la corrección de afinación, sino también la modificación de aspectos tímbricos, dinámicos, articulatorios o incluso identitarios de la interpretación vocal. Desde el sutil *double-tracking* utilizado por John Lennon hasta los sistemas contemporáneos de corrección automática y síntesis vocal, observamos cómo la mediación tecnológica se integra progresivamente en los procesos de construcción de la voz popular grabada. En este contexto, la artificialidad vocal no opera únicamente como efecto expresivo o signo estilístico, sino también como herramienta de adecuación sonora a determinados estándares estéticos e industriales. En términos estrictamente técnicos, buena parte de estas herramientas permiten compensar limitaciones vocales que, en otros contextos históricos, habrían condicionado seriamente la posibilidad de grabar o profesionalizar determinadas interpretaciones.

La corrección de afinación, la homogeneización tímbrica o la manipulación articuladora permiten aproximar ejecuciones vocales dispares a modelos sonoros cultural e industrialmente legitimados. Ciertas competencias tradicionalmente asociadas al dominio técnico del canto pasan a redistribuirse parcialmente entre intérprete, software y producción. Como han señalado autores vinculados a los estudios de producción musical, la grabación contemporánea no puede entenderse ya como simple registro transparente de una interpretación previa, sino como un espacio de construcción sonora donde edición, procesamiento y diseño tímbrico participan activamente en la configuración del resultado final (Moorfield, 2010; Zagorski-Thomas, 2014).

Más que eliminar la necesidad de una performance vocal, estas tecnologías desplazan parte de sus exigencias hacia el entorno técnico de grabación y edición. La voz grabada contemporánea puede construirse progresivamente mediante fragmentación, selección, corrección y procesamiento, haciendo posible que artistas con capacidades vocales limitadas accedan igualmente a modelos de sonoridad competitivos dentro del mercado musical. La cuestión ya no reside únicamente en “cantar bien” en términos tradicionales, sino en integrarse eficazmente en un ecosistema de producción donde interpretación y postproducción aparecen cada vez más interdependientes. De esta manera, la intervención tecnológica puede operar como una forma parcial de sustitución funcional: no reemplaza completamente la presencia del intérprete, pero sí externaliza determinadas capacidades históricamente vinculadas a la técnica vocal.

Del mismo modo que, como apunta Frith (1996), las nociones de autenticidad en la música popular dependen de convenciones culturales e históricas de escucha, también las ideas de “buena” o “mala” voz se encuentran mediadas por estándares técnicos, estéticos e industriales variables según cada contexto. La asistencia técnica aparece así menos vinculada

a la idea de “engaño” que a la adaptación a determinados regímenes de producción y recepción sonora culturalmente normalizados. No hablamos únicamente de asistencia, sino también de formas parciales de delegación vocal. Productores, *beatmakers* o *DJs* pueden hoy construir identidades vocales funcionales sin responder necesariamente al modelo tradicional de cantante virtuoso. Del mismo modo, las tecnologías actuales permiten componer pensando ya en voces transformadas, corregidas o directamente sintetizadas. La voz deja así de entenderse exclusivamente como manifestación orgánica del sujeto que canta para integrarse en procesos de producción tecnomateriales más amplios. La dimensión performativa no desaparece, pero coexiste con nuevas competencias vinculadas a la manipulación, programación y diseño de la vocalidad grabada.

e) Fricción

Finalmente, resulta necesario recordar la importancia del rechazo en la propia evolución del arte. El antropólogo catalán Josep Martí i Pérez (1996; 2000) trabaja con el concepto de patrón de rechazo para describir mecanismos que excluyen determinados productos culturales por motivos estéticos, sociales o identitarios. Aunque su aplicación original se centra en la música y en su legitimación o deslegitimación social, aquí extendemos el alcance del concepto hacia la recepción artística en general, entendiendo el rechazo no sólo como filtro de legitimidad, sino también como motor de transformación. Desde las vanguardias históricas hasta prácticas contemporáneas basadas en la disrupción, la fricción y la provocación, numerosos autores han señalado que la resistencia del público, la crítica o las instituciones puede funcionar como fuerza generativa que impulsa nuevas formas artísticas.¹¹ La historia del arte no se puede concebir sin la fricción entre lo que se considera bello y

¹¹ En esta línea pueden situarse reflexiones anteriores sobre la recepción adversa como condición del arte moderno (Greenberg, 1939), la fricción estética y la resistencia del público como signo de autonomía artística (Adorno, 1970), la redefinición de las fronteras del arte a través del conflicto institucional (Danto, 1964; Dickie, 1974) o las dinámicas de campo y lucha simbólica que transforman el gusto y el canon (Bourdieu, 1992).

lo que no. Y esta fricción, sin la que el arte sería probablemente un terreno mucho más estéril, se extiende a toda una red de condicionantes culturales: desde nuestros modos de enculturación hasta pequeños detalles que, sin que seamos conscientes de ello, decantan la balanza de nuestros juicios sobre el valor de una obra.

Por su parte, Richard Middleton (1990) argumenta que el rechazo a la voz sintética parte de esa condición de elemento orgánico, pues, como se ha comentado anteriormente, es el único instrumento que tenemos tanto los músicos como los no músicos. Sin embargo, hemos visto que esa transgresión que a algunos genera rechazo puede ser altamente atractiva para otros. Según Middleton la fricción entre lo aceptado y lo rechazado en la música no solo refleja los gustos personales, sino que también responde a dinámicas culturales más amplias que determinan qué sonidos se consideran auténticos o válidos en determinados contextos. En el caso de la voz sintética, esta fricción se debe en parte a su capacidad para desafiar las nociones tradicionales de naturalidad, autenticidad y expresión humana. Aquí se abre también la cuestión generacional: lo que a un público le parece artificioso o inaceptable (el Auto-Tune de Cher en 1998), a otro le resulta natural y hasta clásico (el trap o el reguetón en los últimos lustros). Se trata de mediaciones opacas que, con el tiempo, se convertirán en transparentes, naturalizando su recepción a la par que generando nuevos juicios de valor pactados colectivamente. La fricción, por tanto, no es estática: se mueve con las sensibilidades sociales en constante negociación con la herencia cultural y la memoria colectiva.

Coda. Redes Neuronales, *Deep Learning* y IA

Tras décadas de mediación, procesamiento y transformación vocal, la pregunta final resulta casi inevitable: ¿qué ocurre cuando la tecnología deja de interactuar con la voz humana y

pasa directamente a producirla? Si la historia que va del Vocoder al Auto-Tune puede entenderse como un proceso progresivo de tecnificación de la voz cantada, la década actual parece inaugurar una fase distinta: la voz sin cuerpo. Las técnicas contemporáneas de aprendizaje profundo han situado la síntesis vocal en un punto sin precedentes, ya no como simple imitación mecánica, sino como simulación perceptivamente verosímil. Los modelos neuronales de síntesis y clonación vocal pueden reproducir prosodia, fluctuaciones respiratorias, vacilaciones articulatorias e incluso parte de aquello que Barthes habría identificado como el “grano” de la voz. La pregunta ya no es únicamente qué efectos aplicamos sobre una voz humana, sino qué entendemos hoy por “voz” en un contexto donde la separación entre interpretación, simulación y reproducción resulta cada vez más difusa. Sergi Jordà —investigador clave en el ámbito de las interfaces musicales y vinculado al desarrollo de Vocaloid— advertía recientemente que las grandes discográficas comienzan a cuestionar abiertamente el entrenamiento de modelos de IA mediante repertorios y voces protegidas por copyright, apuntando a una inevitable transformación del marco legal contemporáneo. Al mismo tiempo, recordaba que los derechos de autor constituyen un fenómeno históricamente reciente frente a la larga duración antropológica de la música: “La música tiene más de 40.000 años, la industria apenas cien; pase lo que pase, la música seguirá existiendo” (Jordà, 2023). El propio Jordà sugería además que la industria probablemente evolucionaría desde posiciones puramente proteccionistas hacia modelos híbridos de colaboración y licencia con las grandes plataformas de generación musical mediante IA. Los acontecimientos recientes parecen apuntar precisamente en esa dirección: tras las demandas interpuestas en 2024 contra plataformas como Suno o Udio por el uso no autorizado de

grabaciones comerciales para entrenar modelos generativos, varias grandes compañías musicales han comenzado a explorar acuerdos de licencia y colaboración con estas mismas empresas.¹²

Pero más allá del ámbito jurídico y empresarial, lo que está en juego es la concepción misma de la voz. Hasta ahora, incluso la voz procesada mantenía un vínculo corporal reconocible: Lennon tras el ADT, Stevie Wonder abrazando el Talkbox, Cher exhibiendo Auto-Tune, Daft Punk teatralizando el Vocoder. Hoy la voz puede existir sin sujeto identificable, sin biografía y sin fronteras. ¿Qué significa emocionarse con un canto que no pertenece a nadie? ¿Qué implica escuchar una voz que no nació en una garganta? ¿Qué ocurre con el concepto de autoría cuando la voz puede clonarse o generarse sintéticamente? Para una generación que convive con asistentes de voz, audiolibros sintéticos y pop hiperproducido, la voz artificial ya no es necesariamente extraña. Como hemos apuntado anteriormente, la fricción se desplaza: lo que antes era ruptura hoy es norma. El desafío creativo será, quizá, reactivar el conflicto —volver a sorprender, conmover o inquietar con lo artificial normalizado— allí donde la novedad técnica ha dejado de serlo. Desde perspectivas poshumanistas como las de Haraway (1991) o Braidotti (2013; 2019), esta hibridación no es anomalía sino condición cultural del presente, en la que el cuerpo y la subjetividad pueden entenderse como ensamblajes tecnomateriales distribuidos. En este marco, la voz no es ya un índice exclusivo del cuerpo sino un ensamblaje tecnocultural, modulable y reproducible. Las IA no solo continúan la tradición experimental iniciada por Kempelen o Dudley: replantean qué entendemos por identidad vocal, por memoria sonora y por presencia en la escucha. Tal vez —como sugiere Jordà— la música sobreviva a cualquier crisis tecnológica. Pero el modo en que nos relacionemos con las voces, cómo confiamos en ellas,

12 THE GUARDIAN. Warner Music signs deal with AI song generator Suno after settling lawsuit. The Guardian, 26 nov. 2025. Disponible en: <https://www.theguardian.com/business/2025/nov/26/warner-music-signs-deal-with-ai-song-generator-suno-after-settling-lawsuit>. Acceso: 21 mayo 2026

cómo las creemos, cómo las amamos, sí parece condenado a transformarse. Entre fascinación y sospecha, entre nostalgia y utopía, se juega ya el futuro de la voz musical.

Referencias

ADORNO, Theodor W. **Teoría estética**. Madrid: Akal, 2004.

ASKERØI, Eirik. **Reading pop production**: sonic markers and musical identity. 2013. Tesis doctoral. University of Agder, 2013.

BARTHES, Roland. **Image music text**: essays selected and translated by Stephen Heath. London: Fontana Press, 1977.

BELL, Alexander Graham. **Lectures upon the mechanism of speech**. [s.l.]: Alexander Melville Bell & Sons, 1872.

BELL, Alexander Graham. **Upon the formation of a deaf variety of the human race**. Washington: National Academy of Sciences, 1881.

BELL, Alexander Melville. **Visible speech: the science of universal alphabetics**. [s.l.]: Simpkin, Marshall & Co., 1867.

BOURDIEU, Pierre. **Las reglas del arte**: génesis y estructura del campo literario. Barcelona: Anagrama, 1995.

BRAIDOTTI, Rosi. **The posthuman**. Cambridge: Polity Press, 2013.

BRAIDOTTI, Rosi. **Posthuman knowledge**. Cambridge: Polity Press, 2019.

BRØVIG-HANSEN, Ragnhild. Opaque mediation: the cut-and-paste groove in DJ Food's "Break". En: DANIELSEN, Anne (Ed.). **Musical rhythm in the age of digital reproduction**. London: Routledge, 2010. p. 193-210. DOI: <https://doi.org/10.4324/9781315596983-10>.

COWELL, Frederick R. **Life in ancient Rome**. New York: Berklee Publishing, 1980.

DANTO, Arthur C. The artworld. **The Journal of Philosophy**, [s.l.], v. 61, n. 19, p. 571–584, 1964. Disponible en: <https://www.helsinki.fi/assets/drupal/2024-03/Danto-Artworld-1964.pdf>. Acceso: 13 mayo 2026.

DICKIE, George. **The art circle: a theory of art**. New York: Haven Publications, 1984.

DUDLEY, Homer; TARNOCZY, Thomas. The speaking machine of Wolfgang von Kempelen. **The Journal of the Acoustical Society of America**, v. 22, n. 2, p. 151–166, 1950.

FISHER, Mark. **Ghosts of my life: writings on depression, hauntology and lost futures**. Winchester: Zero Books, 2014.

FOOTMAN, Tim. **Welcome to the machine**: OK Computer and the death of the classic album. New Malden: Chrome Dreams, 2007.

FRITH, Simon. Hacia una estética de la música popular. En: CRUCES, Francisco (coord.). **Las culturas musicales: lecturas de etnomusicología**. Madrid: Trotta, 2001. p. 413–436.

GREENBERG, Clement. Avant-garde and kitsch. **Partisan Review**, n. 6, p. 34–49, 1939.

HANDBERG, Kristian. **There is no time like the past**: retro between memory and materiality in contemporary culture. 2014. Tesis doctoral. Facultad de Humanidades, University of Copenhagen, 2014.

HANKINS, Thomas L.; SILVERMAN, Robert J. Vox mechanica: the history of speaking machines. En: HANKINS, Thomas L.; SILVERMAN, Robert J. **Instruments and the imagination**. Princeton: Princeton University Press, 1995. p. 178–220. Disponible en: <http://www.jstor.org/stable/j.ctt7ztvvhb.12>. Acceso: 13 mayo 2026.

HARAWAY, Donna. **Simians, cyborgs, and women**: the reinvention of nature. New York: Routledge, 1991.

HELMHOLTZ, Hermann L. F. **Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik.** Braunschweig: Vieweg, 1863.

HUYSEN, Andreas. *Present pasts: urban palimpsests and the politics of memory.* Stanford: Stanford University Press, 2003.

JORDÀ, Sergi. Veremos cambios radicales, tanto en la manera de crear como de consumir música, debidos a la inteligencia artificial. **Focus UPF**, Universitat Pompeu Fabra, 2023. Disponible en: <https://www.upf.edu/web/focus/w/sergi-jord%C3%A0-veurem-canvis-radicals-tant-en-la-manera-de-crear-com-de-consumir-m%C3%BAsica-deguts-a-la-intel%C2%B7lig%C3%A8ncia-artificial->. Acceso: 13 mayo 2026.

JUAN DE DIOS CUARTAS, M. A.; ROQUER GONZÁLEZ, Jordi. La producción musical: un reto para la musicología del siglo XXI. **Cuadernos de Etnomusicología**, v. 15, n. 2, p. 43–56, 2020. Disponible en: https://www.sibetrans.com/etno/public/docs/5-intro-dossier-produccion_1.pdf. Acceso: 13 mayo 2026.

KEMPELEN, Wolfgang von. **The mechanism of human speech.** Oregon: Wipf and Stock Publishers, 2022.

LACASSE, Serge. **Listen to my voice: the evocative power of vocal staging in recorded rock music and other forms of vocal expression.** 2000. Tesis doctoral. University of Liverpool, 2000.

LINDSAY, David. **Talking head.** Invention & Technology, [s.l.], 1997. Disponible en: <https://www.inventionandtech.com/content/talking-head-1>. Acceso: 13 mayo 2026.

NEIL YOUNG ARCHIVES. "Album of the Week: Trans". 12 jun. 2019, citado en: A Pop Life. Acceso: 16 mayo 2026.

MARTÍ I PÉREZ, Josep. **El folklorismo: usos y abusos de la tradición.** Barcelona: Ronsel, 1996.

MARTÍ I PÉREZ, Josep. **Más allá del arte:** la música como generadora de realidades sociales. Madrid: Alianza, 2000.

McCLARY, Susan. **Feminine endings:** music, gender, and sexuality. Minneapolis: University of Minnesota Press, 2001.

MIDDLETON, Richard. **Studying popular music.** Milton Keynes: Open University Press, 1990.

MOORFIELD, Virgil. **The producer as composer: shaping the sounds of popular music.** Cambridge: The MIT Press, 2010.

MORAN, Caitlin. I Was Feeling Incredible Hysteria and Panic. **Select,** Londres, n. 92, jul. 1997.

NEEDHAM, Joseph. **Science and civilization in China** (volume 4, part 2). Taipei: Caves Books, 1986.

ORD-HUME, Arthur W. J. G. **Automatic pianos:** a collector's guide to the pianola, barrel piano, and Aeolian Orchestrelle. Surrey: Schiffer Publishing, 2004.

SCHROEDER, Manfred. A brief history of synthetic speech. **Speech Communication**, [s.l.], v. 13, p. 231–237, 1993.

SMITH, Jacob. Tearing speech to pieces: voice technologies of the 1940s. **Music, Sound, and the Moving Image**, [s.l.], v. 2, n. 1, p. 77–99, 2008.

TAGG, Philip. **Music's meanings:** a modern musicology for non-musos. New York: Mass Media Music Scholars' Press, 2013.

THE GUARDIAN. Warner Music signs deal with AI song generator Suno after settling lawsuit. **The Guardian**, Londres, 26 nov. 2025. Disponible en: <https://www.theguardian.com/business/2025/nov/26/warner-music-signs-deal-with-ai-song-generator-suno-after-settling-lawsuit>. Acceso: 21 mayo 2026.

TOMPKINS, Dave. **How to wreck a nice beach**: the vocoder from World War II to hip-hop, the machine speaks. London: Stopmusic Books, 2010.

REY, Mauricio. **Tweens, música e Internet**: El consumo musical más allá de la supervisión parental. En E. Encabo (Ed.), *Música y pantallas: cultura, sociedad y educación* (pp. 167–180). Difácil, 2020.

REY, Mauricio. **Tweens y música**: aproximación a las prácticas de consumo musical de los preadolescentes del Área Metropolitana de Barcelona. 2024. Tesis doctoral. Universitat Autònoma de Barcelona, 2024.

REYNOLDS, Simon. **Retromania**: pop culture's addiction to its own past. London: Faber and Faber, 2011.

SHKLOVSKY, Viktor. Art as technique. En: LEMON, Lee T.; REIS, Marion J. (Eds.). **Russian formalist criticism**: four essays. Lincoln: University of Nebraska Press, 1965. p. 3–24. Disponible en: <https://www.literatureoftheamericas.com/wp-content/uploads/2018/08/Art-as-Technique.pdf>. Acceso: 13 mayo 2026.

ZAGORSKI-THOMAS, Simon. **The musicology of record production**. Cambridge:

Cambridge University Press, 2014.

Referencias discográficas

BABBITT, Harry. "I'd Know You Anywhere". **You'll Find Out**. RKO Pictures, 1940.

BON JOVI. "Livin' on a Prayer". **Slippery When Wet**. Mercury Records, 1986.

BON JOVI. "It's My Life". **Crush**. Island Records, 2000.

CARLOS, Wendy. "March from A Clockwork Orange (Ninth Symphony, Fourth Movement)". Stanley Kubrick's A **Clockwork Orange**. Warner Bros. Records, 1972.

CHER. "Believe". **Believe**. Warner Bros., 1998.

DAFT PUNK. "Harder, Better, Faster, Stronger". **Discovery**. Virgin Records, 2001.

DAFT PUNK. "One More Time". **Discovery**. Virgin Records, 2001.

DAFT PUNK. "Digital Love". **Discovery**. Virgin Records, 2001.

DAFT PUNK feat. THE WEEKND. "Starboy". **Starboy**. Republic Records, 2016.

DAFT PUNK feat. THE WEEKND. "I Feel It Coming". **Starboy**. Republic Records, 2016.

DAFT PUNK feat. JULIAN CASABLANCAS. "Instant Crush". **Random Access Memories**. Columbia Records, 2013.

DRAKE, Pete. "Forever". **Forever**. Starday Records, 1964.

ELECTRIC LIGHT ORCHESTRA. "Mr. Blue Sky". **Out of the Blue**. Jet Records, 1977.

EXTREME. "Who Cares?". **Three Sides to Every Story**. A&M Records, 1992.

FRAMPTON, Peter. "Show Me the Way". **Frampton Comes Alive!**. A&M Records, 1976.

FRAMPTON, Peter. "Do You Feel Like We Do?". **Frampton Comes Alive!**. A&M Records, 1976.

KRAFTWERK. "The Robots". **The Man-Machine**. EMI, 1978.

MATHEWS, Max; KELLY, John; GERSTMAN, Louis; LOCHBAUM, Carol.
"Daisy Bell (**Bicycle Built for Two**)". Bell Labs, 1961.

NEWCLEUS. "Jam on It". **Jam on Revenge**. Sunnyview Records, 1984.

RADIOHEAD. "Fitter Happier". **OK Computer**. Parlophone, 1997.

STEVIE WONDER. "Close to You". Actuación en directo en The David Frost Show. 1972.

WALSH, Joe. "Rocky Mountain Way". **The Smoker You Drink, the Player You Get**. ABC-Dunhill Records, 1973.

AFRIKA BAMBAATAA & THE SOULSONIC FORCE. "Planet Rock". **Planet Rock: The Album**. Tommy Boy Records, 1982.

ANDERSON, Laurie. "O Superman". **Big Science**. Warner Bros., 1982.

YOUNG, Neil. **Trans**. Geffen Records, 1982.

Publisher

Universidad Federal de Goiás. Escuela de Música. Programa de Posgrado en Música. Publicación en el Portal de Revistas de la UFG. Las ideas expresadas en este artículo son responsabilidad de sus autores y no representan necesariamente la opinión de los editores ni de la universidad.